# CANADIAN JOURNAL OF PSYCHOLOGY

VOLUME 13, NO. 3 SEPTEMBER, 1959



PUBLISHED FOR THE
CANADIAN PSYCHOLOGICAL ASSOCIATION BY THE
UNIVERSITY OF TORONTO PRESS

AUTHORIZED AS SECOND-CLASS MAIL, POST-OFFICE DEPARTMENT, OTTAWA

# CANADIAN JOURNAL OF PSYCHOLOGY

*Editor*: JULIAN BLACKBURN *Assistant Editor*: P. H. R. JAMES

THE CANADIAN JOURNAL OF PSYCHOLOGY is published quarterly in March, June, September, and December. Annual subscription, $4.00; single number, $1.25.

*Subscriptions.* Orders and correspondence regarding subscriptions, change of address, and purchase of back numbers should be sent to:

THE SECRETARY-TREASURER, CANADIAN PSYCHOLOGICAL ASSOCIATION

Box 31, Postal Station D., Ottawa, Ontario

*Contributions.* Original manuscripts and correspondence on editorial matters should he sent to:

THE EDITOR, CANADIAN JOURNAL OF PSYCHOLOGY

Queen's University, Kingston, Ontario.

*Information for contributors.* The *Journal* publishes experimental and theoretical articles in all recognized fields of psychology. Contributors need not be members of the CPA or residents of Canada. Manuscripts should be submitted *in duplicate*, they should be double-spaced throughout, and they should follow standard practice as regards tables, figures, etc. Authors are asked to adopt the current practice in American and British journals of citing references by date, for example, Smith (1956), instead of by number. "Immediate publication" (i.e. in the next issue to go to press) can be arranged for authors willing to pay the extra costs involved.


## CANADIAN PSYCHOLOGICAL ASSOCIATION, 1959–60

*President*: G. H. TURNER, London; *Past President*: DALBIR BINDRA, Montreal; *President Elect*: R. B. BROMILEY, Toronto; *Secretary-Treasurer*: F. R. WAKE, Ottawa.


The Canadian Psychological Association also publishes *The Canadian Psychologist*, which is distributed to members only. *Editor*: W. R. N. BLAIR, 462 Melbourne Ave., Ottawa, Ont.

# EXPERIMENTAL PSYCHOLOGY AND THE PROBLEM OF BEHAVIOUR DISORDERS[1]

DALBIR BINDRA

*McGill University*

THE PURPOSE of this paper is to suggest a set of principles that might serve as useful guidelines for further research in the area of behaviour disorders or mental illness. These principles are derived mainly from experimental psychology and are opposed to the currently popular "psychodynamic approach." The thesis of this paper is that they provide the only sure way of accumulating reliable knowledge that could serve as a basis for the diagnosis and treatment of behaviour disorders.

The accumulation of reliable and accurate information is, of course, a slow process. However, for better or worse, society's demand for the diagnosis and treatment of man's ills has seldom waited upon the accumulation of scientifically impeccable knowledge. Typically, the professions have developed before the sciences with which they have come to be associated: engineering is older than physics, medicine than physiology and bacteriology, and surgery than anatomy. Thus, it should surprise no one that, although the science of psychology is still in its infancy, there exist two socially recognized professions concerned with the diagnosis and treatment of behaviour disorders. One is psychiatry; the other, clinical psychology. Though the stated aims of and training in the fields of psychiatry, clinical psychology, and experimental psychology are different from each other, many psychiatrists and clinical psychologists are engaged in experimental research. Thus, when I label psychiatrists and clinical psychologists as clinical workers, I shall be referring only to the clinical part of their work; at other times the same persons may function as experimental researchers.

## EVALUATION OF CURRENT STATUS

I shall begin by posing the problem presented by behaviour disorders and evaluating the contribution of the "psychodynamic approach" to the study of their nature, diagnosis, and treatment.

[1]This paper is an adaptation for publication of the presidential address given at the Fifteenth Annual Meeting of the Canadian Psychological Association, Saskatoon, June, 1959. The author gratefully acknowledges the help received from his McGill colleagues in the preparation of the address.

*The Problem*

A person with a behaviour disorder is one whose behaviour is persistently and markedly different from that of the majority of his cultural group, in a way that is considered undesirable by the group or its appointed experts. Whether it arises from a series of anxiety-linked experiences, or from brain impairment, or from abnormalities of body chemistry, the basic identifying feature of any behaviour disorder lies in the frequencies of occurrence of various individual and social activities relative to the frequencies of occurrence of the same activities in a defined relevant group. Clearly, then, the central problem in understanding the causes or psychopathology of a disorder is one of delineating the relevant categories of activities and experimentally analysing the factors that control them.

At the more practical level of dealing with patients, one faces two separate problems. The first of these is that of finding reliable measures obtainable from a small sample of behaviour which, together with other information about the age, sex, past history, employment, present complaint, and so on, would enable one to make with some assurance certain predictive statements about classification, aetiology, therapeutic action, or prognosis, or any combination of these. For convenience, the development of reliable methods for making such predictive statements about a patient may be referred to as the problem of diagnosis. The second practical problem is that of finding, for each type of patient, a course of therapeutic action that will change the patient's frequencies of occurrence of the various classes of activities in the direction of the pattern of activities of the group norm. This is the problem of treatment.

Now, what has the currently popular research approach yielded by way of reliable knowledge about the problems of causation or psychopathology, diagnosis, and treatment of behaviour disorders?

*Agonizing Appraisal*

For the past two or three decades discussions of the problems of causation, diagnosis, and treatment of behaviour disorders have been dominated by what may be called the "psychodynamic approach." It is difficult to give a precise and generally acceptable definition of the psychodynamic approach, but perhaps everyone would agree that it refers to a view of behaviour which stresses "motivation." A large number of conscious and unconscious motivational entities, wishes, desires, tendencies, complexes, frustrations, anxieties, and so on, are considered as the crucial variables determining both normal and abnormal behaviour. These motivational entities are said to develop through the interaction of constitution and experience; however, in practice, experience or learning is

looked upon as the more variable and, therefore, as the more important factor in determining individual differences in the quality, aim, and intensity of the motivational entities. In the area of behaviour disorders, the answer to the problem of psychopathology is said to lie in the "psychodynamics" of the case, and diagnosis is looked upon as the process of discovering the "psychodynamics," that is, of determining the dominant motivations of the patient and the circumstances in which they were acquired. The problem of treatment is then considered as one of altering the patient's behaviour through changing the aims or directions of his motivations.

Anyone who has carefully looked through the last three decades of psychodynamically oriented research knows that it does not provide an occasion for rejoicing and self-congratulation. In fact it seems to have produced findings that, for the most part, either are unreliable (that is, cannot be reproduced) or are of little positive significance.

Consider, for example, the mass of research with the Rorschach and other similar personality tests designed to yield meaningful predictive statements about a person on the basis of an assessment of "his psychodynamics." The negative verdict on the claims of psychodynamically oriented tests arrived at in the reviews of Windle (**33**), and Zubin and Windle (**36**), and in reviews of the literature on personality assessment techniques published in the recent volumes of the *Annual Review of Psychology*, is well known by now. There are also a number of negative findings obtained in carefully controlled recent studies of investigators such as Holtzman and Sells (**12**), Kelly and Fiske (**15**), Kostlan (**16**), and Sines (**29**). In general, these studies show that the increment in predictive accuracy made when the results of the psychodynamic type of psychological tests are added to certain data of an actuarial type—age, education, referral source, etc.—is close to nil; indeed, sometimes the predictions become less accurate when psychodynamical interpretations are added to the actuarial data. Rapid accumulation of negative results of this type has led Meehl (**25**) to challenge anyone to cite a consistent body of published evidence showing that predictions based on "psychodynamics" or "the structure and dynamics of an individual's motivations" are superior to the predictions based on information of an actuarial type. In spite of the great confidence that many clinical workers place in the value of psychodynamics to them, no one to my knowledge has accepted Meehl's challenge.

Turn now to psychotherapy, which is the main procedure of treatment typically followed by clinical workers who subscribe to the psychodynamic approach. The label "psychotherapy" includes a variety of methods involving social interaction and systematic use by the therapist

of any one or more of the techniques labelled catharsis, suggestion, interpretation, and insight. Though strong opinions are held by many concerning the effectiveness of psychotherapy, at present there appears to be no unequivocal evidence that psychotherapy, as defined above, contributes to recovery from behaviour disorders (2). Meehl, who is both a critical scientist and a practising psychotherapist, has summed up the current evidence on psychotherapy in one schizophrenic statement: "Like all therapists, I personally experience an utter inability not to believe I effect results in individual cases; but as a psychologist I know it is foolish to take this conviction at face value" (24, p. 373). Meehl goes on to say, "Our daily therapeutic experiences . . . can be explained within a crude statistical model of the patient-therapist population that assigns very little specific 'power' to therapeutic intervention" (24, pp. 373–374).

By way of conclusion, it may be said that it is not possible to *prove* that the psychodynamic approach will never produce any reliable information or useful techniques; it is conceivable that in the next ten, twenty, or fifty years it may yield something worthwhile. However, in view of the available published research it appears that the confidence which many workers place in the value of the psychodynamic approach in their research and practice is misplaced; the ritual of going through the psychodynamic test and therapeutic procedures may impress the patient, may give the clinical worker a *feeling* of "understanding," and may increase the confidence he places in his own predictions and treatments, but it has no demonstrable clinical or research value. The available research also suggests that the psychodynamic approach, like so many other ideas in the history of science, has turned out to be a wrong "lead." Thus, it seems to me that any further work along this approach would not constitute the most efficient strategy of research.

## Some Guidelines for Future Research

In the remainder of this paper I shall attempt to indicate a strategy of research which might prove more fruitful than the psychodynamic approach. I shall do no more than attempt to make explicit and bring together those principles in psychological and medical research that are particularly relevant to the study of behaviour disorders.

*Principle 1*

*Research on the problems of causation, diagnosis, and treatment of behaviour disorders should concentrate, not on "psychodynamics" or other hypothetical processes, but on observed behaviour.*

Typically, a clinical report consists of two parts. In the first part the clinical worker describes the behaviour of the patient as he has observed

it or as it has been described to him. He follows this with certain facts about the present circumstances of the patient and information about the family and personal histories. These data consist of such items as age, socio-economic level, performance at school, the symptoms as observed, and circumstances of onset of the disorder. All these items of information can be, and usually are, determined in a fairly objective and reliable way. In the second part of the clinical report the worker reports the results of some personality tests of questionable reliability and validity, and then adds his over-all impression of the "psychodynamics" of the case. Statements in this part of the report may read as "Marked anxiety reaction in an immature individual with passive dependent needs" (Garfield, **10**, p. 225), or "Strong pent-up aggressions are indicated and, although the aggressions appear to be absorbed to a large extent in fantasy, the poor controls implied in his impulsiveness probably permit aggressive, anti-social outbursts" (Schafer, **27**, p. 232). Such statements obviously lack any precise, operational meaning. Thus, the clinical worker often moves from reliable data in the first part of his report to vague, interpretative statements about hypothetical processes in the second part.

No matter how vague and untidy these interpretative statements may be, they do serve an important function for the clinical worker. They provide him with some sort of a rationale for deciding upon a course of therapeutic action, and, in the absence of exact knowledge of the aetiology of most behaviour disorders, there is no other basis for making the administrative decisions as to which one of several treatments to give a patient. Thus, so far as the objectives of clinical work are concerned, it may be desirable to postulate hypothetical psychodynamic processes as temporary and tentative substitutes for exact knowledge of aetiology. However, whatever the requirements of practical, administrative decisions, there can be no doubt that for purposes of research on psychopathology, and for developing improved methods of diagnosis and treatment, one must concentrate on behaviour as observed. For, in moving from observed behaviour to the type of hypothetical processes usually postulated, one abandons the more reliable in favour of the less reliable; error and vagueness are introduced. And this is no way to develop reliable knowledge. Therefore, as far as possible, we should minimize the use of hypothetical entities, and link behavioural data directly to other empirical variables, such as the course of a disorder, environmental and physiological factors that produce fluctuations in symptoms, and effects of particular treatments. The practical implication of this principle is that those clinicians who also function as research workers must work with different sets of concepts and must step outside the framework of

psychodynamic theory and practice for improving their diagnostic and therapeutic techniques for the future.

*Principle 2*

*Descriptions of subjective states, not being subject to publicly observable or objective reliability checks, should not be considered as statements about crucial psychopathological events; however, verbal statements about such states may, under certain conditions, serve as reliable data.*

Subjective states such as those of pain, anxiety, depression, elation, triumph, jealousy, helplessness, and dejection are reported by almost all human beings, and, in a common sense sort of way, it appears reasonable that personal descriptions of such states be considered as evidence in discussions of psychopathology. One often hears that the diagnostician or the therapist should aim at determining not what the patient does but what he "feels," not how the investigator manipulates the patient's environment but what the change "means" to the patient, not what the patient is but what he "perceives" himself to be. Now, as everyone knows, scientific evidence must be open to objective reliability checks. Therefore, no personal descriptions of subjective states can be employed as evidence in discussions of hypotheses concerning psychopathology, no matter how "real" such states may be to the patients describing them.

It is hardly necessary to review at length the utter confusion that resulted from the introspectionist schools of Titchener and Külpe in their attempts to answer psychological questions by regarding descriptions of subjective states as crucial evidence. Two sentences from Boring's paper (4) on the history of introspection tell the story: "Classical introspection, it seems to me, went out of style after Titchener's death [1927] because it had demonstrated no functional use and therefore seemed dull, and also because it was unreliable. Laboratory atmosphere crept into the descriptions, and it was not possible to verify, from one laboratory to another, the introspective accounts of the consciousnesses of action, feeling, choice, and judgment." (4, p. 174.) In view of this historical lesson it is regrettable that many workers in the field of psychopathology continue to consider descriptions of subjective states as evidence, and often formulate diagnostic and therapeutic questions in terms of such states.

It should be noted that the above statement does not deny the possibility of the use of verbal statements about subjective states as data in their own right. Indeed, it is quite legitimate to consider statements such as "I am depressed" and "I feel anxious" as dependent variables and to investigate the frequency of occurrence of these responses under dif-

ferent experimental conditions. An analysis of the conditions which determine the occurrence of such verbal responses would be a contribution to the study of linguistic behaviour, and might also provide the knowledge necessary to link meaningfully verbal data with causal, diagnostic, and therapeutic considerations in the area of behaviour disorders. However, these types of correlations can be established only by treating verbal statements as data rather than as crucial evidence, and completely ignoring the subjective connotations of the statements.

*Principle 3*

*The aim of diagnostic and research testing procedures should be the measurement of significant aspects of behaviour: that is, tests should measure variables whose relations with other dependent and independent variables in general psychology have been well established experimentally rather than* ad hoc *variables which appear temporarily to be of some practical significance.*

By a significant psychological variable I mean nothing more than a variable which has been studied extensively and has been shown to be related to a variety of other behavioural, social, or biological variables. Any variable that is meaningfully related to a large number of other empirical variables may be considered as a more significant variable than one that is not so related. (This criterion for determining whether a variable can be considered "fundamental" is closer to Margenau's (**23**) concept of basic validity than to Cronbach and Meehl's (**6**) and Loevinger's (**19**) concept of construct validity.)

Recent experimental work in both animal and human laboratories has delineated many significant variables that are of direct relevance to the problems of behaviour disorders, that is, they appear to be meaningfully related to the variable of presence *vs.* absence of particular symptoms. For instance, Malmo and Shagass (**21**) have shown that normals, neurotics, and psychotics differ from each other in the increase in muscle tension brought about by noxious stimulation; furthermore, they have shown (**22**) a clear relation between a patient's symptoms and the pattern of increases in muscle tension in different parts of the body. Other recently delineated significant variables, which can be objectively and reliably measured and which are relevant to behaviour disorders, include relative response specificity (Lacey, **17**), sedation threshold (Shagass, **28**), susceptibility to arousal and avoidance tendency (Kamin *et al.*, **14**), rate of operant responding (Skinner *et al.*, **30**; Lindsley, **18**), time estimation (Stern, **31**), suggestibility and persistence (Eysenck, **7**), and rate of conditioning (Taylor, **32**). Fundamental variables of this type are already providing a basis for conducting reliable and significant in-

vestigations and for developing diagnostic procedures for use in clinical work. It should be noted that research of this type is initially aimed not at developing tests for practical purposes but at delineating fundamental aspects of behaviour through detailed experimental analysis of relevant phenomena. The laboratory is the best place for this type of analysis.

The fact that the currently popular personality tests fail significantly to increase the validity of diagnostic decisions is partly attributable to their failure to sample fundamental aspects of behaviour: they bear no established relation to the main body of psychological knowledge. Thus, it is hard to see why, for example, a small "d" response on the Rorschach, or agreement with the statement "I find it hard to keep my mind on a task or job" on the Taylor Scale of anxiety, should have any general diagnostic capacity. The fact that these responses are given more frequently by neurotics than by normals may be practically useful information, but it is not clear how these responses are related to other variables of greater empirical and theoretical significance. Since the relation of such responses to a criterion variable remains an isolated fact, it is likely that even the practical utility of this relation lacks cross-situational power, being dependent upon the operation of some adventitious factors peculiar to a particular time and place. Thus, if in 1954 disagreement with the item "I prefer a bath to a shower" was a characteristic response of successful graduate students of the University of California at Berkeley, it is unlikely that the same item would differentiate between successful and unsuccessful graduate students at the University of Edinburgh, or that it would differentiate between these groups today, five years later, even at Berkeley. What is worse is that such test items may cease to correlate with the validating criterion with variations of time and place without the investigator being any the wiser.

In view of these considerations, it is important to concentrate on developing reliable, objective measures of fundamental aspects of behaviour rather than on developing tests which may, for some unknown reason, temporarily appear to be useful. Once some reliable measures of significant variables have been found, the measures are bound to be of significance for both research and clinical purposes. This suggestion is supported by a study of the diagnostic and research tests employed in the field of general medicine. Such widely used medical tests as body temperature, pulse rate, blood pressure, and blood count are all tests that measure fundamental physiological aspects of the body, aspects that are equally relevant to the description of the physiology of normal and sick persons. Take body temperature, for instance. Though there had been some interest in the use of the thermometer in medical diagnosis, it was not until the middle of the last century that variations in body tempera-

ture came to be regarded as an index of a fundamental physiological property of the organism (*see* Mettler, **26**, p. 312). Thus, it was only after an instrument for reliably measuring body temperature was available, and only after the mean body temperature of the normal man was established, that it was possible to do the kind of validation study that laid the foundation of modern clinical thermometry. The pioneering validation study was carried out by Wunderlich (*see* Mettler, **26**) who, after studying hundreds of typhoid patients during an epidemic was able to give an affirmative answer to the question of the existence of a significant relation between the course of the disease and variations in body temperature. I think it is high time that in the field of behaviour disorders it was also recognized that only the fundamental behavioural properties of the organism constitute a reasonable basis for developing tests for clinical and other applied psychological purposes.

*Principle 4*

*The crucial psychological problem in understanding behaviour disorders is that of determining the laws which govern the interaction between habit strength and other factors that control the occurrence of responses.*

It must be recognized once and for all that the behaviour on the basis of which we classify a person as, for example, a neurotic or an organic psychotic is not determined solely by a particular set of anxiety-related experiences or by a particular type of brain damage. As has been pointed out by so many, the functional-organic dichotomy of behaviour disorders does not correspond to facts of psychopathology. Not all those who have undergone anxiety-producing experiences become neurotics, nor do all with a certain type of brain damage show identical behavioural deviations. In the case of a compulsive neurotic, for example, though the specific compulsions could undoubtedly be shown to be related to the life experiences of the patient, a psychologist must still face the problem that not all persons with similar experiences develop the same compulsions, and, indeed, many do not develop any compulsions at all. Similarly, the various activities that characterize the general paretic do not result from brain damage alone; there is no specific type of brain damage that will cause a person to propose marriage to the ward nurse three times a day, or to walk the hospital wards wearing nothing except a hat and a cigar. Of course, since the existence of tertiary syphilitic infection in general paretics has been established, general paresis has come to be defined in terms of the syphilitic damage to brain cells. However, this happy finding, no matter how useful it may be practically, leaves unanswered the psychological question of why the general paretic

behaves as he does. What I am saying is that explaining neurotic or psychotic behaviour involves more than linking it to some aetiological factor, such as particular antecedent experiences or the abnormality of brain function. The medical concept of aetiology is too narrow to cover all aspects of scientific inquiry. In analysing and explaining behaviour disorders, the psychologist faces many subtle problems which go beyond the interests of the practising clinicians.

The essential problem in developing explanations of behaviour disorders seems to reside in the interaction between habit strength, on the one hand, and, on the other, the so-called aetiological factors, be they chemical, organic, or experiential in origin. I employ the term "habit strength" to refer to nothing more than the degree of prepotence acquired by a particular activity. The variable of habit strength occupies a special place among the factors that control behaviour, for, as the habit strength of an activity increases, it seems to become functionally autonomous, or relatively independent of the chemical, situational, and other factors that initially controlled it (see 3). Thus, what I consider to be the crucial interaction is the one between the habit strengths of the various activities that exist in an individual's repertoire and the operation of the so-called aetiological factors such as anxiety-producing experiences, brain damage, and changes in body chemistry. This type of interaction may be illustrated with reference to the recent studies with the psychotomimetic drug, adrenochrome (11).

Suppose that the presence of a certain amount of adrenochrome, or some related substance, is a necessary factor in making a person behave in a way that would lead to his classification as a schizophrenic. Now, as has been pointed out by Hoffer *et al.* (11), the exact effects on behaviour of any experimentally administered adrenochrome are typically not uniform from individual to individual. This variability seems to persist even when an equivalent amount is injected into every subject, and when the behaviour is observed after approximately the same post-injection interval. Some of the reported variability is undoubtedly due to individual differences in the reactivity of the relevant tissues to the drug, but another equally important source of variability seems to lie in the initial differences between subjects in the habit strength of the relevant responses. Thus, assuming adrenochrome to be the necessary factor in schizophrenia, one must ask how the specific activities shown by different schizophrenics are related to the relative habit strengths of the activities existing in their repertoires at the time of the disease onset (i.e. operation of the "aetiological factor"), as well as to the new responses acquired after the onset. Specifically, the psychologist must learn (*a*) the exact way in which adrenochrome affects the occur-

rence of activities of varying habit strength that already exist in a person's repertoire, (*b*) how adrenochrome affects the acquisition of new activities, and (*c*) what the behavioural characteristics of high adrenochrome individuals reared in different environmental ("cultural") settings are. Knowledge of such interaction between habit strength and adrenochrome will tell us not only what the exact effects of a given dosage of adrenochrome will be, but also how, through appropriate combinations of drugs and psychological training, one could minimize the occurrence of given (undesirable) types of activities.

I believe that some fundamental experimental research designed to determine the general laws which govern the interaction between habit strength on the one hand and factors such as body chemistry, arousal, and sensory cues on the other is the most important single problem facing psychologists who are interested in understanding the phenomena of behaviour disorders. Any systematic therapeutic use of drugs or psychological retraining, or a combination of the two, presupposes a knowledge of the laws that govern these interactions.

*Principle 5*

*Research on treatment of behaviour disorders should concentrate on developing techniques of manipulating the conditions that currently control the patient's undesirable responses rather than on unearthing the conditions which initially produced his disorder.*

This principle represents a finding that has been repeatedly corroborated in a variety of areas of research. It is what Allport (1) referred to as "functional autonomy of motives," and what, in my recent book (3), I have rechristened as "partial autonomy of activities." In that book I have collated considerable evidence supporting the generalization that with increasing practice an activity becomes relatively independent of the conditions under which that activity was initially acquired or practised. For example, the alteration of certain sensory cues will disrupt the normal sexual responses of a sexually inexperienced male rat, but will not affect the sexual performance of a sexually sophisticated (i.e., high habit-strength) rat. Similarly, the greater the habit strength of a dominance or aggressive activity the more likely is it to withstand variations in gonadal hormones and in other blood factors. The same relation also holds between habit strength and the level of arousal of the organism: the greater the habit strength of an activity the greater appears to be the range of variation of the organism's level of arousal within which that activity will occur without disruption. Clearly, through repeated performance, activities tend to become relatively independent of the conditions which were initially necessary for their performance. This

means that the chemical, situational, and reinforcement factors that maintain an activity after it has been well practised may be different from those that were crucial in its acquisition and early performance.

The phenomenon of partial autonomy of activities implies that in order to eliminate or to decrease the frequency of occurrence of an undesirable activity, it is more important to know and to manipulate the conditions that currently control that activity than to unearth the factors under which it was first acquired. The success of some recent attempts at purely symptomatic treatment (for example, Jones, **13**; Malmo *et al.*, **20**; Wolpe, **34**; Yates, **35**) indicate that Freud (**9**, pp. 253–254) was too hasty, and wrong, in concluding that unearthing the developmental conditions of a neurotic disorder (hysteria) was necessary for its successful treatment. Of course, knowledge of the historical factors which led to the development of an undesirable activity may be sought for its own sake, and such knowledge will undoubtedly increase our understanding of behaviour disorders; however, so far as the practical aim of modifying behaviour is concerned, the factors that currently control it are the more, if not the only, important ones.

These considerations point to one of the reasons for the doubtful efficacy of psychotherapy. Typically, one of the major undertakings in the psychotherapeutic situation is what is called "interpretation." To a large extent interpretation involves linking the patient's symptoms to the factors which did operate or may have operated at the onset of his behaviour disorder. Now, it is certainly doubtful whether it is possible to unearth the factors which actually determined the onset of the symptoms through the social interaction of a psychotherapeutic situation. But even if it is possible, this knowledge on the part of either the therapist or the patient would have little bearing on any attempt to discover and manipulate the variables that currently control the disordered aspects of the patient's behaviour. Perhaps the best course of action would be, first, to subject the patient to some semi-experimental situations to determine the factors that currently control the undesirable activities, and then to consider which factors, situational, chemical, or experiential, or which combination of these, are likely to constitute the most effective treatment. "Interpretation" and "insight," in the sense in which these terms are employed in psychotherapy, have little bearing on this procedure.

A recent study by Coons (**5**) is relevant to this point. Coons compared the effects on hospital patients of two types of group psychotherapy. In one type the technique stressed interaction among members of the group without reference to personal difficulties of individual patients. In the other condition the technique stressed "cognitive understanding of per-

sonal difficulties (insight)." Coons found that the interaction group showed significantly greater improvement in adjustment than did the insight group; the insight group did not differ from a control group which had received no treatment at all. Thus, Coons concludes that interaction rather than insight seems to be the essential condition for therapeutic change. The failure of insight by itself to produce any favourable change is consistent with the proposition that getting to know the factors that initially produced a disorder may have little bearing upon treatment of that disorder. The efficacy of the interaction procedure may be attributed to the fact that the patients were made to engage in the very types of activities which are considered as representing improved adjustment.

The work of Wolpe (34) and Ferster (8) also bears on this issue. Both have shown how experimental findings in the field of learning may be used as a basis for modifying the behaviour of patients without getting involved in interpretation, insight, or other similar procedures employed in traditional psychotherapy. Some would claim that traditional psychotherapy also seeks to employ the principles of learning but does so at a verbal, symbolic level. But the fact that psychotherapy has not been shown to constitute an effective treatment of behaviour disorders means that the current psychodynamically oriented techniques of re-education (catharsis, interpretation, insight, etc.) apparently fail to make use of those principles of learning which might constitute a sound basis for effecting change in behaviour. Indeed, the procedures suggested by Wolpe and Ferster are so different from those currently subsumed under the rubric of "psychotherapy" that they should probably be given a separate name—perhaps "retraining" or "response re-education."

To recapitulate, treatment of behaviour disorders must necessarily involve replacing undesirable activities with desirable ones. This can be done by subjecting the patient to the type of procedures—and these include situational, chemical, surgical and all other types of factors that control response—that would increase the probability of occurrence of the desired activities. Such procedures must involve the active manipulation of real, palpable conditions that currently control his behaviour, rather than merely the indirect, symbolic re-education that may take place in the conventional psychotherapist's office.

## Summary and Conclusion

Three decades of psychodynamically oriented research, which stresses conscious and unconscious wishes, desires, frustrations, anxieties, and other motivational entities as the determiners of normal and abnormal behaviour, have failed to contribute significantly to the problems of

causation or psychopathology, diagnosis, and treatment of behaviour disorders or mental illness. This paper suggests a set of principles that are likely to serve as useful alternative guidelines for further research in the area of behaviour disorders. The principles are derived mainly from the research experience of experimental psychologists. The proposed principles are:

1. Research on the problems of causation, diagnosis, and treatment of behaviour disorders should concentrate, not on "psychodynamics" or other hypothetical processes, but on observed behaviour.

2. Descriptions of subjective states, not being subject to publicly observable or objective reliability checks, should not be considered as statements about crucial psychopathological events; however, verbal statements about such states may, under certain conditions, serve as reliable data.

3. The aim of diagnostic and research testing procedures should be the measurement of significant aspects of behaviour: that is, tests should measure variables whose relations with other dependent and independent variables in general psychology have been well established experimentally rather than *ad hoc* variables which appear temporarily to be of some practical significance.

4. The crucial psychological problem in understanding behaviour disorders is that of determining the laws which govern the interaction between habit strength and other factors that control the occurrence of responses.

5. Research on treatment of behaviour disorders should concentrate on developing techniques of manipulating the conditions that currently control the patient's undesirable responses rather than on unearthing the conditions which initially produced his disorder.

Only if these principles are followed in research is there any hope of accumulating a systematic body of reliable knowledge about behaviour disorders, knowledge which would be not only scientifically sound but also practically useful. It is to be hoped that more and more mental hospitals will start laboratories for experimental research and that more and more experimental psychologists will interest themselves in the field of behaviour disorders. If this happens, there can be no doubt that psychology will, in the near future, provide the practice of psychiatry and clinical psychology with the same type of reliable scientific foundation as physiology has provided for the practice of general medicine.

## REFERENCES

1. Allport, G. W. *Personality.* New York: Holt, 1937.
2. Bindra, D. Psychotherapy and the recovery from neurosis. *J. abnorm. soc. Psychol.*, 1956, **53**, 251–254.
3. Bindra, D. *Motivation: A systematic reinterpretation.* New York: Ronald, 1959.
4. Boring, E. G. A history of introspection. *Psychol. Bull.*, 1953, **50**, 169–189.
5. Coons, W. H. Interaction and insight in group psychotherapy. *Canad. J. Psychol.*, 1957, **11**, 1–8.

6. Cronbach, L. J., & Meehl, P. L. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.
7. Eysenck, H. J. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952.
8. Ferster, C. B. Reinforcement and punishment in the control of human behavior by social agencies. *Psychiatric Research Reports*, 1958, **10**, 101–118.
9. Freud, S. *Collected papers*, vol. I. London: Hogarth Press and Institute of Psycho-analysis, 1953.
10. Garfield, S. L. *Introductory clinical psychology*. New York: Macmillan, 1957.
11. Hoffer, A., Osmond, H., & Smythies, J. Schizophrenia: A new approach. II. Result of a year's research. *J. ment. Sci.*, 1954, **100**, 29–45.
12. Holtzman, W. H., & Sells, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.
13. Jones, H. Gwynne. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–419.
14. Kamin, L. J., Bindra, D., Clark, J. W., & Waksberg, Helene. The interrelations among some behavioural measures of anxiety. *Canad. J. Psychol.*, 1955, **9**, 79–83.
15. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor, Mich.: Univer. Mich. Press, 1951.
16. Kostlan, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, **18**, 83–88.
17. Lacey, J. I., Bateman, Dorothy E., & Van Lehn, Ruth. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15**, 8–21.
18. Lindsley, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiatric Research Reports*, 1956, **5**, 118–139.
19. Loevinger, Jane. Objective tests as instruments of psychological theory. *Psychol. Reports, Monogr. Supplement* 9, 1957, **3**, 635–694.
20. Malmo, R. B., Davis, J. F., & Barza, S. Total hysterical deafness: An experimental case study. *J. Pers.*, 1952, **21**, 188–204.
21. Malmo, R. B., & Shagass, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24.
22. Malmo, R. B., & Shagass, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11**, 25–29.
23. Margenau, H. *The nature of physical reality*. New York: McGraw-Hill, 1950.
24. Meehl, P. E. Psychotherapy. Chap. 14 in *Annual Review of Psychology*, 1955, **6**, 357–378.
25. Meehl, P. E. Some ruminations on the validation of clinical procedures. *Canad. J. Psychol.*, 1959, **13**, 102–126.
26. Mettler, Cecilia C. *History of medicine*. Philadelphia: Blakiston, 1947.
27. Schafer, R. *The clinical application of psychological tests*. New York: Internat. Universities Press, 1948.
28. Shagass, C. Sedation threshold: A neurophysiological tool for psychosomatic research. *Psychosom. Med.*, 1956, **18**, 410–419.
29. Sines, L. K. An experimental investigation of the relative contribution to clinical diagnosis and personality description of various kinds of pertinent data. Unpublished Ph.D. thesis, Univer. Minnesota, 1957.
30. Skinner, B. F., Solomon, H. C., & Lindsley, O. R. A new method for the experimental analysis of the behavior of psychotic patients. *J. nerv. ment. Dis.*, 1954, **120**, 403–406.

31. Stern, Muriel H. Thyroid function and activity, speed, and timing aspects of behaviour. *Canad. J. Psychol.*, 1959, **13**, 43–48.
32. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–90.
33. Windle, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451–482.
34. Wolpe, J. *Psychotherapy by reciprocal inhibition.* Stanford, Calif.: Stanford Univer. Press, 1958.
35. Yates, A. J. The application of learning theory to the treatment of tics. *J. abnorm. soc. Psychol.*, 1958, **56**, 175–182.
36. Zubin, J., & Windle, C. Psychological prognosis of outcome in the mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272–281.

# SPONTANEOUS ALTERNATION AFTER FREE AND FORCED TRIALS[1]

WILLIAM N. DEMBER AND HARRY FOWLER
*Yale University*

THERE ARE two methods typically employed in research on spontaneous alternation behaviour. In one the animal is allowed two successive free trials; in the other the animal's first-trial response is "forced," in the sense that only one alternative is made available, and the second trial is free.

In reviewing the literature on spontaneous alternation (2) we noted that studies using the forced-trial method generally showed a higher percentage of alternation than those using the free-trial method. This procedural difference was cited as one possible source of the discrepancies among data collected by different investigators.

The conclusion that the forced-trial method yields a higher percentage of alternation than the free-trial method is supported only by comparisons across experiments. No single study has been reported in which the two methods were directly compared. In this paper, two such comparisons are presented, and in addition some supplementary data are provided on the relation between percentage of alternation and length of the inter-trial interval.

## METHOD

### *Subjects*

The subjects were 64 male albino rats, approximately 100 days old. Each S was housed individually in a wire mesh cage and maintained on an *ad libitum* diet of Purina chow and water. The Ss were handled 10 min. per day for 5 days prior to the experiment.

### *Apparatus*

Alternation testing took place in a T-maze which was composed of a starting compartment 8 in. long, a starting alley 12 in. long, and 2 goal arms, each 20 in. long. Alley widths were all 4 in., and alley walls were 4¾ in. high. The floor and walls of the starting compartment and alley were painted grey; the left goal arm was black, the right arm white. A grey guillotine door separated the starting compartment from the starting alley; similar doors, one black and the other white, separated the goal arms from the starting alley.

### *Procedure*

Two experiments were conducted with 32 Ss in each. In both experiments every S participated in each of the 4 conditions created by a 2 × 2 design of free- *vs.*

[1]Supported by a grant from the National Science Foundation.

forced-trial procedure and short *vs.* long inter-trial interval. In Experiment I, inter-trial intervals were 2 and 16 min.; in Experiment II, the intervals were 2 and 32 min. Each S was run in one condition on each of 4 days, with the test days separated by 48 hrs. Assignment of an S to a condition on a particular test day was made by reference to a Latin-square design.

On a given day each S had two trials in the T-maze. The first trial was either free or forced. A free-trial S always ran before a forced-trial S. Forcing was accomplished by lowering the door of one arm prior to the S's introduction into the maze. For a given forced-trial S, the open arm was the one entered by the preceding free-trial S. In this way, the forced-trial condition provided the same number of trial-one entries into each arm as the free-trial condition.

After the S had entered an arm on the first trial, the door was lowered behind it. The S was removed from the arm after about 10 sec. and returned to its home cage. The S was put back into the maze and given a free trial after 2, 16, or 32 min.

## Results

Data are presented in Table I relevant to the free *vs.* forced comparison and to the comparison between the short- and long-interval conditions. A percentage of alternation significantly different from 50 per cent ($p < .05$, sign test) was obtained in every condition except the free-trial conditions of Experiment II. However, in both experiments, there was a preference for the black arm, 62.5 per cent in Experiment I and 68 per cent in Experiment II, making 50 per cent an overestimate of the chance level of alternation. If preference is taken into consideration, the chance level of alternation in Experiment II is 43.5 per cent; the obtained percentage in the 2-min. condition differs significantly from the chance level ($p < .05$, $t$-test), but this is not true of the 32-min. condition.

For every interval condition, the forced-trial method produced more alternation than the free-trial method. Over all intervals, the forced condition yielded 81.2 per cent alternation, the free condition 64.8 per cent.

TABLE I

Per cent Alternation in Each Experimental Condition

| Experiment | Interval (min.) | Trial procedure | | Over-all percentage |
|---|---|---|---|---|
| | | Free | Forced | |
| I | 2 | 75.0 | 87.5 | 81.2 |
| | 16 | 65.6 | 78.1 | 71.8 |
| II | 2 | 62.5 | 81.3 | 71.9 |
| | 32 | 56.3 | 78.1 | 67.2 |
| Over-all percentage | | 64.8 | 81.2 | |

The difference between the two conditions is reliable[2] in both Experiment I and II ($p = .05$ and $< .01$, respectively). Summed over both experiments, the difference between free and forced conditions is significant beyond the .001 level of confidence.

For every trial condition, the short inter-trial interval produced more alternation than the long interval. Summed over trial conditions and over both experiments, the short interval yielded 76.5 per cent alternation, the long interval 69.5 per cent. Though the difference between short- and long-interval conditions is not significant within each experiment, over both experiments the difference is reliable ($p = .05$). Because of the difference in the results of the 2 min. interval condition between experiments I and II, it is not possible to compare meaningfully the difference in amount of alternation between the 16 and 32 min. interval conditions.

## Discussion

The difference in the amount of alternation obtained with the two methods confirms our previous observation (2) and lends support to our speculation concerning discrepancies among data in the alternation literature. On theoretical grounds there are at least three reasons for expecting the forced-trial method to yield the higher percentage of alternation.

First, in the free-trial situation the subject, because of its typical vacillation at the choice-point, can partially sample the unchosen alternative, and hence build up some satiation for it (see 2). This experience, which is precluded or minimized in the forced-trial situation, may reduce the alternation tendency in the free-trial situation over that in the forced-trial condition.

Second, the free-trial method is open to a statistical bias: ordinarily, the chance level of alternation is assumed to be 50 per cent. This assumption leads to an overestimate of the chance level if there is a strong initial preference for one alternative; the percentage of alternation obtained may, therefore, be underestimated in the free-trial situation. The forced-trial procedure is less subject to this statistical artifact since trial-one responses are usually distributed equally between the two alternatives.

Finally, the forced-trial procedure introduces an added source of attractiveness to the non-chosen arm in the form of stimulus change: on trial-one the non-chosen arm is blocked, whereas on trial-two it is not. A change of this sort should enhance the subject's tendency to investigate the previously unentered arm (1), and hence increase the apparent amount of alternation over that in the free-trial situation.

[2]The significance of this and of the following differences was assessed by the McNemar test, as described in Siegel (3).

It is not possible to say from the present results which of the suggested bases of the effect of forced trials is the most important. The importance of the statistical artifact has, however, been minimized in the present experiment by equating the distributions of free- and forced-trial entries.

The data indicate further that although the alternation tendency declines as inter-trial interval increases, it is maintained over intervals greater than two minutes. This result adds one more to the list of studies (see 2) showing alternation over "long" inter-trial intervals.

## Summary

In 2 separate experiments, employing 32 rats each, the hypothesis was confirmed that there is more alternation following a forced trial to one arm of a T-maze than following a free trial. The data also revealed more alternation with a short, 2-min., inter-trial interval than with a long, 16- or 32-min. interval. The long intervals did, however, yield significant amounts of alternation, with the exception of the 32-min., free-trial condition. The results of these experiments are consistent both with previous observations and with theoretical expectations.

## REFERENCES

1. Dember, W. N. Response by the rat to environmental change. *J. comp. physiol. Psychol.*, 1956, **49**, 93–95.
2. Dember, W. N.. & Fowler, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.
3. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

# THE EFFECT OF PRACTICE UPON THE PERCEPTION OF CAUSALITY[1]

PETER F. POWESLAND[2]
*Queen's University*

MICHOTTE (8) HAS MADE a thorough investigation of the immediate spatial and temporal stimulus conditions which give rise to the impression of causality, and other experimenters have followed up his work with studies of a similar nature (3, 9, 10, 11). But the phenomenological approach adopted by all these workers has led them to overlook the possible influence of past experience upon judgments made by a subject when asked to say whether or not two events appear to be causally related. It is with this question that the present study is concerned.

An investigation into the effects of past experience upon the perception of causality has been made by Gruber *et al.* (5) in experiments in which the stimulus-event was the collapse of a model "bridge" following the removal of one of its vertical supports. The apparatus was so designed that the interval between the removal of the support and the collapse of the bridge could be varied by the experimenters. As would be predicted from Michotte's results, Gruber found that when this time interval was sufficiently short "a very striking impression of causality" was reported by the subject. Gruber calculated for each subject a "temporal threshold of causality," defined as the maximum delay between the two parts of the stimulus-event at which the subject reliably reported an impression of causality. He examined the effects of interpolating a series of special trials between two assessments of this threshold, and found that a series of trials in which there were long delays between the two parts of the stimulus-event was followed by an increase in the threshold value, while an interpolated series of trials with zero delays tended to be followed by a reduction in the threshold value.

The first of the two experiments reported here was designed with a view to confirming Gruber's results, and to investigating further the

[1]This article is a shortened version of a thesis approved by Queen's University at Kingston, Ontario, as partial fulfilment of the requirements for the degree of Master of Arts. The greater part of the research reported was carried out during the author's tenure of an R. Samuel McLaughlin Resident Fellowship at Queen's University. Additional financial support received through an award by the National Research Council of Canada is also acknowledged.

[2]The author is now at the Department of Psychology, University of Bristol, England.

influence of a prior series of trials upon the temporal threshold of causality.

Gruber's treatment of the problem was extended in two ways. First, since he had used the same number of interpolated trials throughout his experiment, it was decided to examine the effects of varying the length of the interpolated series. Second, it was decided to compare the effect of an interpolated series of trials, in which the subject was required to judge each stimulus-event, with a series in which he merely observed but made no overt responses. In a preliminary report of his findings (4) Gruber noted that during the interpolated series of trials with long delays many subjects "complained of boredom, fatigue, or discomfort at having to give so many 'not causal' responses." This suggests that the changes in judgment following the interpolated trials may have been at least partially related to the act of responding itself. The possible importance of overt responses is also indicated by studies which have shown that verbal responses to stimuli assist retention and recall (1, 7). Although no overt recall of the interpolated stimuli was required in Gruber's experiment, it is possible that, as in the retention experiments, verbal responses established traces of the past stimuli more firmly and clearly, thus improving their later efficacy. If this is true, a series of interpolated trials to which the subject does not respond should prove to have a reduced effect upon the threshold of causality. A further indication of the importance of responding is given in an experiment by Brown (2) in which it was found that the effect of an interpolated stimulus was reduced when the subject was not required to judge it.

The second of the two experiments reported here was designed to examine the effect of a prior series of trials upon the threshold of causality measured with respect to Michotte's *l'effet lancement*. The effect of an interpolated series of trials in which there was a long delay between the two parts of the stimulus-event was compared with the effect of a series in which there was no delay, and the effect of interpolated trials to which the subject responded was compared with the effect of interpolated trials to which he did not respond.

## Experiment I

The purpose of this experiment was to examine the effects of a series of practice trials interpolated between two assessments of the temporal threshold of causality, in terms of the following hypotheses: (1) that the temporal threshold of causality will be raised as a consequence of an interpolated series of trials in which there is a long delay between the two parts of the stimulus-event; (2) that the temporal threshold of causality will be lowered as a consequence of an interpolated series of

trials in which there is no delay between the two parts of the stimulus-event; (3) that the effect of an interpolated series of trials upon the temporal threshold of causality will increase as the number of interpolated trials is increased; (4) that the effect of an interpolated series of trials upon the temporal threshold of causality will be greater when S makes overt judgments of the interpolated trials than when he does not.

*Apparatus*

The apparatus was a slightly modified version of Gruber's "bridge," and consisted of two upright wooden posts surmounted by a horizontal wooden bar 27 cm. long. The wooden bar was hinged to the left-hand upright post and could be held in a horizontal position by the attraction between an electromagnet concealed within the post and a small metal plate inlaid in the bar. The right-hand post was mounted on a wheeled platform which was attached to the cores of two solenoids in such a way that the energizing of one solenoid held the upright immediately under the right-hand end of the bar and the energizing of the other pulled it away to the right. These solenoids were connected through electrical relays to a telegraph key. Depression of the telegraph key caused the right-hand upright to move swiftly away to the right. At the precise moment when the rear edge of the movable post passed from under the right-hand end of the horizontal bar, a relay was automatically tripped, actuating an electronic interval-timer. The timer kept the concealed magnet energized for a predetermined period and then released it, thus allowing the horizontal bar to fall.

In front of the apparatus was placed a large screen in which there was a window 33 cm. long and 22 cm. high. This allowed S to see only the two uprights, approximately 12 cm. high, and the horizontal bar. When the right-hand post was moved to the right, it came to rest outside S's field of vision. Lighting was provided by two 25-watt bulbs placed behind the screen, one on each side of the window, at a level slightly above that of the horizontal bar. The apparatus was backed by a large sheet of white cardboard.

A small electric bell was attached to the apparatus, for use as a warning signal at the beginning of each trial.

*Subjects*

Forty voluntary subjects (24 women and 16 men) were drawn from the academic and administrative staffs and the student body of the university. They were divided into 8 groups in the manner described in the following section.

*Experimental Design and Procedure*

Ss were tested individually in a darkened room, and were seated 2 m. in front of the screen. The apparatus and procedure were described in the following way:

"This bar rests on the left-hand post and is held there by a magnet. I control the magnet, so I determine when the bar will fall. Now I have a switch, back here, and when I press this switch (*E* presses telegraph key, with interval-timer set at 5 sec.), the right-hand post moves away to the right, but as you can see, the bar can remain in a horizontal position until I release it. (Bar falls.) Now, under certain conditions, when I move the post and let the bar fall, it may *look as if* the bar falls *because* the post has moved away. (*E* demonstrates with zero interval.) Actually, there is no causal relation between the movement of the post and the falling of the

bar, because the bar is controlled by the magnet. But what I want you to do during this experiment is to say—each time, when I move the post and release the bar—whether or not it *looks as if* the removal of the post is what causes the bar to fall. I would like you to say "yes" if you think it looks as if the bar falls because the post has been moved away, and "no" if you don't think so. Each time before I move the post I will ring this bell [rings] as a warning signal. Now let's start with some practice trials to give you the idea of how the experiment works."

*Preliminary practice trials.* The preliminary practice period consisted of 10 trials using the following delays between the removal of the right-hand post and the falling of the bar, in the order shown: (seconds) 0.40, 0.00, 1.00, 0.10, 0.20, 0.02, 2.00, 0.01, 0.80, 0.05.

*The pretest.* The method of single stimuli was adapted in the following manner to measure the temporal threshold of causality for each S. A series of 14 delays, varying in steps of 0.01 sec. was presented to S 4 times, a different random arrangement of the series being used each time. The threshold was thus derived from 56 judgments by S. The random order of presentation was different for each S. The range of delays to be used for a particular S was chosen on the basis of his responses during the preliminary practice period. In most cases, a range of delays from 0.00 sec. through 0.13 sec. was appropriate, but for 4 of the Ss the range from 0.07 sec. through 0.20 sec. was used. The threshold was computed by Spearman's arithmetical method.

*The interpolated treatment period.* Following immediately after the pretest, and before a second assessment of the temporal threshold of causality in the post-test, Ss were given an interpolated "treatment" period. The 8 groups of Ss were distinguished according to the nature of this treatment period as shown below. Five Ss were allocated at random to each group.

Group C1: 30 trials in which the delays were of the same order of magnitude as those used in the pretest, S making an overt response after each trial (30 responding normal-range trials)

Group C2: a 10-min. rest period (during which S conversed with the experimenter on matters not connected with the experiment)

Group E1: 50 trials using zero delays, S observing but making no overt response (50 non-responding zero trials)

Group E2: 50 trials using zero delays, S making an overt response after each trial (50 responding zero trials)

Group E3: 10 trials using long delays,[3] S making an overt response after each trial (10 responding delay trials)

Group E4: 30 trials using long delays,[3] S making an overt response after each trial (30 responding delay trials)

Group E5: 50 trials using long delays,[3] S observing but making no overt response (50 non-responding delay trials)

Group E6: 50 trials using long delays,[3] S making an overt response after each trial (50 responding delay trials)

*The post-test.* Immediately after the treatment period, the temporal threshold of causality was again assessed for each S, by the method used in the pretest. When necessary the range of delays was extended upwards in order to arrive at a threshold value.

---

[3]The "long delays" were of the following values, presented in random order: (seconds) 0.40, 0.50, 0.60, 0.70, 0.80.

## Results

Threshold values obtained for each subject in the pretest and the post-test were compared. Table I shows the median changes in threshold

TABLE I

Experiment I: Changes in Temporal Threshold of Causality following Interpolated Treatments

| Group | Treatment | Median pretest threshold (milliseconds) | Median post-test threshold (milliseconds) | Median change | Significance of change |
|---|---|---|---|---|---|
| C1 | 30 responding normal-range trials | 70.0 | 60.0 | −12.5 | Not significant |
| C2 | Rest period | 67.5 | 67.5 | 0.0 | Not significant |
| E1 | 50 non-responding zero trials | 77.5 | 52.5 | −12.5 | Not significant |
| E2 | 50 responding zero trials | 75.0 | 75.0 | −15.0 | Not significant |
| E3 | 10 responding delay trials | 55.0 | 45.0 | −15.0 | Not significant |
| E4 | 30 responding delay trials | 55.0 | 75.0 | +7.5 | Not significant |
| E5 | 50 non-responding delay trials | 57.5 | 77.5 | +15.0 | $p = 0.031$ (one-tailed) |
| E6 | 50 responding delay trials | 60.0 | 80.0 | +10.0 | $p = 0.031$ (one-tailed) |

values for each group. The complete set of raw data obtained in the course of the experiment was subjected to the Kruskal-Wallis One-Way Analysis of Variance and the over-all contingency was judged to be significant ($\chi_r^2 = 14.01$; $df = 7$; $p \simeq 0.05$).

The pretest and post-test data for each group were examined by means of the Walsh Test and it was found that only in Groups E5 (50 non-responding delay trials) and E6 (50 responding delay trials) were the distributions of threshold changes significantly different from distributions which could be expected to occur by chance ($p = 0.031$ for a one-tailed test, in both cases). As predicted, the threshold changes in these two groups tended to be positive (see Table I).

The Mann-Whitney U Test was used to compare the effects of the different treatments administered to the various groups. This analysis showed that:

(*a*) There was no significant difference between the effects of the two control treatments (C1 and C2).

(*b*) The effect of practice with zero delays (E1, E2) did not differ significantly from the effects of treatment in the control groups (C1, C2).

(*c*) Both groups using 50 delay trials (E5, E6) showed threshold changes significantly different from those found in the control group which rested between the pretest and the post-test (C2) ($p < 0.004$ in both cases).

(*d*) The effect of practice with 50 non-responding delay trials (E5) was significantly different ($p < 0.048$) from the effect of 30 normal-range trials (C1), but the effect of practice with 50 responding delay trials (E6) was not ($p = 0.075$).

(*e*) Groups E5 and E6 both showed changes in threshold significantly different from those in the two control groups combined (C1 and C2) ($p < 0.01$ in both cases).

(*f*) There was no significant difference ($p = 0.075$) between the effect of 50 non-responding zero trials (E1) and the effect of 50 non-responding delay trials (E5), but the effect of 50 responding zero trials (E2) differed significantly ($p < 0.028$) from the effect of 50 responding delay trials (E6).

(*g*) When the effect of practice with 50 delay trials (Groups E1 and E2 combined) was compared with the effect of practice with 50 zero trials (Groups E5 and E6 combined), disregarding the factor of responding/non-responding, the difference was significant ($p < 0.01$).

(*h*) The effects of 10 delay trials (E3) and of 30 delay trials (E4) did not differ significantly from the effects of the control treatments, or from each other.

(*i*) The effect of 10 responding delay trials (E3) and of 30 responding delay trials (E4) did not differ significantly from the effects of practice with zero trials (E1, E2).

(*j*) The effect of 10 responding delay trials (E3) differed significantly ($p = 0.028$) from the effect of 50 non-responding delay trials (E5) but not from the effect of 50 responding delay trials (E6).

(*k*) The effect of 30 responding delay trials (E4) was not significantly different from the effects of 50 non-responding or of 50 responding delay trials (E5, E6).

(*l*) Comparison between Groups E1 and E2, and between E5 and E6 indicated that there was no significant difference between the effect of practice during which S responded and the effect of equivalent practice during which he did not respond.

These results support Hypothesis 1 in that they show a tendency for practice with long delays to be followed by a raising of the temporal threshold of causality, but there is no satisfactory evidence in favour of

Hypothesis 2. The finding that the use of 50 interpolated delay trials produced a significant change in the threshold value, whereas the use of 10 and 30 interpolated delay trials did not, may be regarded as evidence tending to confirm Hypothesis 3. There is no evidence to support Hypothesis 4.

## Experiment II

The purpose of this experiment was to examine the effects of a series of practice trials interpolated between two assessments of the temporal threshold of causality, in terms of the following hypotheses: (1) that the temporal threshold of causality will be raised as a consequence of an interpolated series of trials in which there is a long delay between the two parts of the stimulus-event; (2) that the temporal threshold of causality will be lowered as a consequence of an interpolated series of trials in which there is no delay between the two parts of the stimulus-event; (3) that the effect of an interpolated series of trials upon the temporal threshold of causality will be greater when S makes overt judgments of the interpolated trials than when he does not.

### *Apparatus*

The basic design of the apparatus used by Michotte in *la méthode des disques* was followed (8). In this method a large white vertical disc is rotated behind an aperture, 15 cm. long and 0.5 cm. high, cut in a screen which lies flush with the surface of the disc. An arc of a circle drawn on the disc so as to be concentric with the axis of rotation appears in the viewing slit as a small stationary object; a segment of a spiral curve falling towards the centre of the disc appears as a small moving object. The speed and timing of the apparent movement of an object depend upon the characteristics of the drawing and upon the rate at which the disc is rotated (8, p. 26 ff.). With a suitable arrangement of lines, the following kinematic sequence is made to appear: (1) A small black square (object A) appears at the left-hand end of the horizontal slit, and a small red square (object B) appears in the centre. (2) After a pause, object A moves towards object B, makes contact with it and thereafter remains stationary. (3) Either immediately or after an interval, object B moves to the right and disappears from view. (4) The slit becomes empty for a short time and then the sequence of events is repeated. When the ratio between the speeds of the two objects is suitable and when the time interval separating their movements is sufficiently short, A appears to strike B and push it away to the right. This is the phenomenon called by Michotte *'l'effet lancement'*.

For the present experiments, object A moved at a speed of 36 cm./sec. and object B moved at a speed of 10.08 cm./sec. (ratio, 3.6:1). These speeds were chosen, on the basis of Michotte's findings, as being favourable to *l'effet lancement*.

By the use of two concentric discs—one for object A and one for object B— it was possible to vary the time interval between the end of A's movement and the beginning of B's movement by altering the angular displacement of the one disc relative to the other.

The discs were rotated by means of an electric motor, coupled with suitable reduction gear.

The testing room was illuminated by overhead fluorescent lighting, and a 100-watt reading lamp was focused on the aperture of the screen.

*Subjects*

Forty-eight voluntary subjects (26 women and 22 men) were drawn from the academic and administrative staffs and the student body of the university. They were divided into 5 groups, as will be described later.

*Experimental Design and Procedure*

Ss were tested individually, seated 2 m. in front of the screen. At the beginning of the session, S was asked to watch the slit in the screen, where something would appear. He was told to concentrate on the centre of the slit. The electric motor was then switched on and the stimulus sequence, with object B moving off as soon as it was struck by object A, was allowed to repeat itself 5 times. When the motor was switched off (after 5 complete revolutions of the disc) S was asked to describe what he had seen. In the majority of cases, S's reply was to the effect that the black square had struck the red square and pushed it away to the right. In some cases, however, *l'effet lancement* was not immediately apparent, and object A was said to pass object B, or to change places with it. In these cases the stimulus-event was shown again after E had given a verbal description of the movements of the two objects and instructed S to look for a relationship between them. Usually a description amounting to *l'effet lancement* was eventually elicited, but one S was unable to see any relationship between the movements of the two objects, and she was replaced.

S was told that during the course of the experiment he would be shown a large number of events similar to the one he had just seen, and that in some of these it might look as if the movement of object A caused the movement of object B, but that in some the two movements might seem to be independent of each other. He was asked to say "yes" or "no" in each case, according to whether or not it looked to him as if object A was pushing object B to the right. It was explained that each stimulus-event would be shown 3 times in immediate succession and that he was to make his judgment after the third viewing.[4] S was told that the experiment would begin with a series of practice trials.

*Preliminary practice trials.* A series of practice trials was then given, using the following angular displacements of the two discs, in the order shown:[5] (degrees) 00, 60, 80, 40, 20, 30, 50, 10, 70. Each of these delays was used 3 times in immediate succession, S responding after the third viewing of each.

---

[4]This particular method of presenting the stimulus-event was adopted in order to minimize any undesirable effects associated with the starting and stopping of the electric motor. The procedure rendered negligible any slight deviations from the correct operating speed by spreading them over two presentations of the stimulus-event.

[5]The formula for converting the scale of degrees of arc to one of time intervals is:

$$t = \frac{D}{6R}$$

where $t$ = the time interval (in seconds) between the end of A's movement and the beginning of B's movement

$D$ = the angular displacement (in degrees of arc) of the one disc relative to the other

$R$ = the angular velocity of the discs (in r.p.m.)

Since, in the present experiment, the discs were rotated at 24 r.p.m. a displacement of 1 degree of arc yielded a time interval of approximately 0.007 sec.

*The pretest.* The method of single stimuli was adapted in the following manner to measure the temporal threshold of causality for each S, in terms of the delay between the impact of object A upon object B and the beginning of B's movement to the right. Twenty different delays were listed in random order, using a scale of degrees of arc to represent time intervals. The delays varied in steps of 2 degrees of arc (i.e., approximately 0.014 sec. of time). Each stimulus-event was presented 3 times in immediate succession and S made his response of "yes" or "no" after the third viewing. Each trio of identical stimulus-events was presented once only. The range of delays to be used for each S was chosen on the basis of his responses during the preliminary practice period. In most cases the range from 2 degrees through 40 degrees of arc was appropriate, but in 3 cases a somewhat higher range was used. The threshold was defined as the mid-point of the zone of uncertainty.

*The interpolated treatment period.* Following immediately after the pretest, and before a second assessment of the temporal threshold of causality in the post-test, Ss were given an interpolated "treatment" period. The 5 groups of Ss were distinguished according to the nature of this treatment period, as shown below.

Group C (8 Ss): Rest period (during which S conversed with the experimenter on matters not connected with the experiment)

Group E1 (10 Ss): 60 trials with long (0.80 sec.) delays, the presentations being in 20 groups of 3, S making a judgment after each trio (20 × 3 responding delay trials)

Group E2 (10 Ss): 60 trials with zero delays, the presentations being in 20 groups of 3, S making a judgment after each trio (20 × 3 responding zero trials)

Group E3 (10 Ss): 60 trials with long (0.80) delays, the presentations being in 20 groups of 3, with a pause, but no judgment by S, after each trio (20 × 3 non-responding delay trials)

Group E4 (10 Ss): 60 trials with zero delays, the presentations being in 20 groups of 3, with a pause, but no judgment by S, after each trio (20 × 3 non-responding zero trials)

*The post-test.* Immediately after the treatment period the temporal threshold of causality was again assessed for each S by the method used in the pretest. When necessary the range of delays was extended upwards in order to arrive at the threshold value.

## Results

Threshold values obtained for each subject in the pretest and the post-test were compared. Table II shows the median values of the threshold changes for each group (expressed in the relative units of degrees of arc). The complete set of raw data from which Table II was compiled was subjected to the Kruskal-Wallis One-Way Analysis of Variance, and the over-all contingency was judged to be significant ($\chi_r^2 = 24.86$; df = 4; $p < 0.001$).

For each group the pretest data and the post-test data were examined by means of the Walsh Test and it was found that in Groups E1 (20 × 3 responding delay trials) and E2 (20 × 3 responding zero trials) the distributions of threshold changes were significantly different from dis-

TABLE II

EXPERIMENT II: CHANGES IN TEMPORAL THRESHOLD OF CAUSALITY FOLLOWING INTERPOLATED TREATMENTS

| Group | Treatment | Median pretest threshold (degrees of arc) | Median post-test threshold (degrees of arc) | Median change | Significance of change |
|---|---|---|---|---|---|
| C | Rest period | 19.5 | 22.5 | +0.5 | Not significant |
| E1 | 20×3 responding delay trials | 19.0 | 26.5 | +5.5 | $p < 0.005$ (one-tailed) |
| E2 | 20×3 responding zero trials | 20.5 | 14.0 | −6.0 | $p < 0.005$ (one-tailed) |
| E3 | 20×3 non-responding delay trials | 17.0 | 17.0 | 0.0 | Not significant |
| E4 | 20×3 non-responding zero trials | 22.5 | 19.5 | −1.5 | Not significant |

tributions which could be expected to occur by chance. These changes represented a tendency for the threshold to be raised following an interpolated series of "delay" trials to which the subject responded, and to be lowered following an interpolated series of "zero" trials to which the subject responded (see Table II). In the control group (C), in which no interpolated practice was given, and in the groups in which the subject made no responses during the interpolated practice period (E3 and E4), the distributions of threshold changes were not significantly different from chance distributions.

The Mann-Whitney U Test was used to compare the effects of the different treatments administered to the five groups. This analysis showed that interpolated practice during which the subject made overt responses (E1, E2) produced threshold changes significantly different from those found when he rested between the pretest and the post-test (C) ($p < 0.01$ in each case). The directions of these threshold changes were as predicted by the hypotheses, i.e. an increase in threshold value following practice with long delays, and a reduction in threshold value following practice with no delays (see Table II). Interpolated practice during which the subject made no overt responses (E3, E4) failed to produce changes significantly different from those found in the control group (C). This was not predicted. The hypotheses did predict, however, that interpolated practice during which the subject made overt responses would be more effective than practice during which he did not make overt responses, and this was confirmed by the results of the experiment. Comparison between "responding" and "non-responding" treatment for the

"delay" groups (E1, E3) showed a significant difference in effect ($p < 0.025$). Comparison between "responding" and "non-responding" treatments for the "zero" groups (E2, E4) also showed a significant difference ($p < 0.01$). There was no significant difference between the effects of interpolated "delay" trials and the effects of interpolated "zero" trials, when the subject was not required to respond.

## Discussion

Gruber *et al.* interpreted their results in terms of "an anchoring effect of past trials, modifying the temporal conditions necessary to elicit an impression of causality." The results of the present study are also considered to be sufficiently similar to the findings of previous experiments on the anchoring of judgment scales to warrant the conclusion that similar processes are involved.

From both experiments it is clear that the introduction of a high anchor, in the form of interpolated trials with long delays between the two parts of the stimulus-event, had the effect of raising the value of the temporal threshold of causality. The use of a low anchor, in the form of practice with zero delays, tended to lower the value of the temporal threshold of causality, but in Experiment I the effect was not statistically significant. A somewhat weaker effect of practice with zero delays would be consistent with previous findings that, in general, an anchor stimulus close in value to the original range of stimuli to be judged is less effective than one further removed from the original range. In Experiment I the range of delays used in the pretest lay between 0.00 sec. and 0.13 sec. and the "delay" trials used intervals ranging between 0.40 sec. and 0.80 sec. In Experiment II, the range of delays used in the pretest period lay between 0.00 sec. and approximately 0.28 sec. (0–40 degrees of arc) and the "delay" trials used an interval of 0.80 sec. (115.2 degrees of arc). In both cases, therefore, the high anchor was much further removed from the pretest range of stimuli than was the low anchor of 0.00 sec. delay.

The question of the relative effectiveness of varying numbers of interpolated practice trials was examined only in Experiment I, and the scope of the investigation was limited to comparing the effects of different numbers of interpolated delay trials during all of which the subject responded. There is evidence of a relationship between the number of interpolated practice trials and the anchoring effect produced, in that the use of 50 interpolated trials produced a significant change in the threshold value, whereas the use of 10 and 30 interpolated trials did not. The tentative conclusion that there is a correlation between the amount of change in threshold and the number of presentations of the anchor

stimulus would be in accord with Helson's theory that the adaptation level (represented here by the temporal threshold of causality) results from a "pooling" and averaging of all the stimuli experienced (**6**). If Helson's theory is valid, the efficacy of a particular stimulus in bringing about a significant change of adaptation level would be some function of both its magnitude and the frequency of its occurrence.

The experiments provide apparently conflicting evidence as to the importance of overt responses by the subject during the interpolated practice trials. In Experiment II, significant changes in the threshold values were obtained only when the subject was required to judge the anchor stimuli. In Experiment I, on the other hand, the results did not seem to be dependent upon whether or not he made overt judgments during the interpolated practice periods. There is a possible explanation of this disagreement if it can be assumed that the part played by the activity of responding to an anchor stimulus is to concentrate the subject's attention upon it. In the present experiments two different lighting conditions were used, and it may well be that the level of attention maintained by the subject when viewing but not responding to the interpolated practice trials was affected by these conditions. Experiment I was carried out in a darkened room, and only the apparatus was illuminated. For Experiment II the whole room was illuminated by fluorescent lighting. It is possible that under the former condition the subject was less tempted to allow his gaze and his attention to wander during the periods when he did not have to make overt responses. A further factor which may have affected his level of attention during the "non-responding" trials was the difference in size of the stimulus-objects used in the two experiments. Since the two moving objects used in Experiment II were only 5 mm. square, and were viewed from a distance of 2 m., it is probable that sustained attention to them called for more effort than was required for viewing the much larger "bridge" of Experiment I. If this is so, the subject may have taken the opportunity to relax attention during the "non-responding" trials of Experiment II. A possible, but somewhat speculative, explanation of the disagreement between the two experiments as to the importance of responding during the interpolated trials is, therefore, that in Experiment I attention during the interpolated practice period was maintained whether the subject was responding or not, whereas in Experiment II the level of attention dropped when he did not have to respond.

The findings of Gruber *et al.* (**5**) receive support from the present experiments, but it is of interest that the amounts of the changes in the temporal threshold of causality found in Experiment I were in general considerably less than those found in Gruber's experiment. Gruber found threshold changes of the order of 75 ms., using interpolated delay trials,

and 60 ms. with interpolated zero trials. The threshold changes found in Experiment I of the present study were of the order of 15 ms. This difference may be related to the fact that Gruber used a somewhat coarser scale of stimulus-values in the pretest than that used in the present study. In his experiment, the delays used in the calculation of the threshold varied in steps of 0.05 sec. In Experiment I, reported here, the pretest stimuli varied in steps of 0.01 sec. Perhaps the result of using a finer scale of stimulus-values during the pretest was to train the subject to make finer discriminations between the lengths of the delays, thereby reducing proportionately the magnitude of the threshold change following practice.

The effects of practice with the delay trials and with the zero trials may be interpreted in terms of contrast. It is well known that a sound of medium intensity seems to be relatively quiet when contrasted with one of higher intensity, and relatively loud when contrasted with a sound of lower intensity. Similar effects are produced in other sense modalities. In the present experiments, the long delays and the zero delays may have given rise to an analogous effect and thus changed the subject's concept of what constituted a "short" delay. From incidental remarks made by the subjects in the course of the experiments it was evident that most of them were actively judging the length of the delays between the two parts of the stimulus-events rather than simply allowing the "impression of causality" to make itself felt. Remarks such as "It's difficult to tell with some of these—they're so much alike," and "No. There was a definite pause there" and "Could you show me again the one that was a 'yes'?" all indicate that the subjects tended to treat their task as one of active discrimination rather than of reporting the "immediate impressions" of which Michotte speaks. This being the case, it is far from unlikely that their judgments were influenced by contrast with preceding practice trials. This is not an alternative explanation to that of anchoring, for contrast may be regarded as a special case of anchoring.

## Summary

Two experiments dealing with the effects of practice upon the perception of causality are reported. In Experiment I the stimulus-event consisted of the collapse of a model "bridge" following the removal of one of its vertical supports. The delay between the two parts of the stimulus-event was variable and the "temporal threshold of causality" was defined as the maximum value of this delay consistent with reliable reports by S of an impression of causality. In Experiment II the stimulus event was the apparent impact of one small object upon another. The delay between the impact and the movement of the object struck was variable and the temporal threshold of causality was defined as the maximum value of this delay consistent with reliable reports by S of an impression of causality.

The temporal threshold of causality was measured before and after a series of practice trials, and it was found that practice during which there were long delays between the two parts of the stimulus-event tended to increase the threshold value, and practice during which there were no delays tended to reduce the threshold value.

In Experiment I the evidence suggested that the effect of the interpolated practice increased as the number of practice trials was increased.

The two experiments gave contradictory evidence as to the importance of overt responses by S during the interpolated practice period. In Experiment I the effect of the interpolated practice appeared to be independent of whether or not S judged the practice trials; in Experiment II, significant changes in threshold value were obtained only when S was required to make overt responses during the interpolated practice periods.

Changes in the temporal threshold of causality are interpreted in terms of the anchoring effect of the practice trials interpolated between the pretest and the post-test.

## REFERENCES

1. Bartlett, F. C. *Remembering*. New York: Macmillan, 1932.
2. Brown, D. R. Stimulus-similarity and the anchoring of subjective scales. *Amer. J. Psychol.*, 1953, **66**, 199–214.
3. Burke, L. On the tunnel effect. *Quart. J. exp. Psychol.*, 1952, **4**, 128–139.
4. Gruber, H. E., & Fink, C. D. The role of knowledge and experience in the perception of causality. (Unpublished)
5. Gruber, H. E., Fink, C. D., & Damm, V. Effects of experience on perception of causality. *J. exp. Psychol.*, 1957, **53**, 89–93.
6. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
7. Kurtz, K. H., & Hovland, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.
8. Michotte, A. *La Perception de la causalité*. Louvain: Publications Universitaires de Louvain, 2nd ed., 1954.
9. Olum, V. Developmental differences in the perception of causality. *Amer. J. Psychol.*, 1956, **69**, 417–423.
10. Yela, M. Phenomenal causation at a distance. *Quart. J. exp. Psychol.*, 1952, **4**, 139–154.
11. Yela, M. La nature du "rayon d'action" dans l'impression de causalité mécanique. *J. Psychol. norm. Path.*, 1954, **47–51**, 330–348.

# EFFECT OF DEGREE OF ILLUMINATION ON RATE OF AMBIGUOUS FIGURE REVERSAL[1, 2]

D. CIPYWNYK

*University Hospital, Saskatoon*

THIS PAPER DEALS with the first in a series of experiments designed to investigate sources of variation affecting the reversal-of-perspective phenomenon. The work described here deals with the effect of illumination.

Despite the fact that the mechanism of the reversal-of-perspective phenomenon is not well understood at present, many attempts have been made to use it as an objective test. Among the probable reasons for its persistent use are a high degree of reliability (**10, 11**) and the fact that it appears to be related to certain personality traits (**6, 14**), as well as to changes in the human organism brought about by disease or pharmaceutical agents (**7, 9, 19, 20**). It has also been shown that patients diagnosed as schizophrenic respond differently on this test (**6, 12, 18**). Since the author plans to use this perceptual measure for studying certain aspects of schizophrenic thinking, as well as for assessing the effects of drugs on schizophrenic and normal subjects, other sources of variation which affect this measure should be more thoroughly understood.

Though the effect of illumination has been studied to some extent, the experimental evidence is somewhat contradictory. Mull, Armstrong, and Telfer (**16**) and Mull, Ord, and Locke (**17**) have shown that neither degree of contrast nor amount of illumination has any appreciable effect on the reversal rate. On the other hand, if it is assumed, as it is by Köhler and Wallach (**13, 14**) and Brown (**2, 3**), that satiation is the cause of apparent changes in an ambiguous figure, then the rate of reversal should increase as illumination is increased. The rationale for this prediction is that the increased flow of stimulation per unit time to a given area of the visual cortex would bring about a more rapid satiation for a given aspect of the figure. Support for this view comes from Brown's experiment (**2**) in which he found that the rate of apparent change as a

[1]Supported by the Department of Public Health, Saskatchewan; National Health Grants, Ottawa: and the Rockefeller Foundation. Approved by the Saskatchewan Committee on Schizophrenia Research.

[2]The author is indebted to Miss A. Duthie for her help with the statistical design and calculations. Thanks are also due to Dr. N. Agnew and Mr. A. Levey for invaluable suggestions.

function of time was significantly higher for the binocular than for the monocular condition. It is quite well established that when corresponding (or non-corresponding) parts of the two eyes are stimulated by the same luminance level the resulting brightness is greater than that obtained by stimulation of only one eye (1).

The hypothesis proposed in this study is that the rate of reversal of a Necker cube will be an increasing function of illumination.

## Method

*Apparatus*

The apparatus consisted of a specially constructed viewing box where such factors as visual angle, size of stimulus, and angle of regard could be easily controlled. It was believed that the viewing box would also decrease distraction and ensure attention to the stimulus.

For the present study, the Necker cube was chosen mainly for its simplicity and reported high reliability (10). The dimensions of the cube were as follows: face 1 in. square, offset ¾ in. toward upper right aspect; lines were 1 mm. thick. The cube was centred on a white paper sheet supported by a square plate within the viewing box, and situated approximately 16 in. from S's eyes. This plate was evenly illuminated by a frosted bulb situated in the forward and lower aspect of the box. Changes in illumination were effected by a voltage-regulator. Reversals were recorded by a magnetic counter as S tapped a key each time a change of aspect occurred.

*Subjects and Procedure*

The subjects consisted of 12 female introductory psychology students, with an age range of 19 to 30 years. All reported to be naïve with respect to the reversible perspective phenomenon.

Prior to the test, the two aspects of the cube were pointed out to each S and she was allowed to view an experimental stimulus cube until at least one reversal occurred. She was told that she would view a similar cube in the viewing box and was asked to press the key every time she saw a reversal. Each S was instructed to keep her eyes focused on a small cross in the middle of the cube when the light was turned on.

After this preliminary instruction, S was seated in front of the viewing box and, when ready, the stopwatch and the light illuminating the cube were switched on simultaneously. The stopwatch was reset to zero after the first reversal occurred.

Three degrees of illumination were used: Dim, Medium, and Bright. The voltage drop across a 60-watt bulb as measured with a vacuum-tube voltmeter was 34, 76, and 110 respectively. These three conditions were presented in all possible permutations, that is, 3! ways. Two Ss received the same permutation. For example, Ss 1 and 7 received the permutation Dim, Medium, Bright; Ss 2 and 8 received the permutation Bright, Dim, Medium, etc. The Ss viewed the cube for 2 min. under each condition with a rest interval of 10 min. between each condition. This 10-min. interval was inserted to eliminate any carry-over or residual effects of one treatment upon the following one.

The number of reversals was recorded for each 10-sec. period during the 2-min. performance.

## Results

The statistical analyses carried out on this data were based on Cochran, Autrey, and Cannon (4), Cochran and Cox (5), and Federer (8). Analysis of variance was carried out for performance periods of 30, 60, 90, and 120 seconds. Table I illustrates the analysis carried out at the thirty-second performance period.

TABLE I

Analysis of Variance on Degree of Illumination at the 30-second Performance Period

| Source | df | SS | | MS | | F |
|---|---|---|---|---|---|---|
| Between groups | 3 | 563.2 | | 187.7 | | |
| Between columns within groups | 8 | 4174.0 | | 521.8 | | |
| Between rows within groups | 8 | 537.3 | | 67.2 | | |
| Between treatments | 2 | 259.3 | 162 | 81.0 | | 4.71* |
| Treatment × group interaction | 6 | | 97.3 | 16.2 | | |
| Error | 8 | 143.4 | 240.7 | 17.9 | 17.2 | |
| TOTAL | 35 | 5677.2 | | | | |

*$p < .05$

The design and statistical analysis of a short-term switchover trial, comparing three treatments, was used. Advantages in employing this type of analysis lie in the accuracy of comparisons of the effects of treatments, and in the unbiased estimates of experimental errors.

Four groups, or four 3 × 3 Latin squares, were formed, the columns of which consisted of the 3! permutations of three treatments, arranged in such a manner that no treatment appeared twice in the same row.

The standard analysis of each Latin square was carried out. In Table I, the sums of squares are summations of sums of squares from the individual analysis, with the exception, of course, of the total sum of squares, and the between-groups sum of squares.

Since the differences between the treatments were the same in all groups (apart from experimental errors)—that is, the mean square for this term was no larger than the error mean square—this value was included in the error term, with a corresponding 14 degrees of freedom.

The $F$ values for the 60, 90, and 120-second performance periods were 3.75, 3.88, and 4.67 respectively. All the values are significant at the 5 per cent level.

Table II illustrates the mean number of reversals at the 30, 60, 90,

TABLE II

MEAN NUMBER OF REVERSALS FOR EACH 30 SECONDS OF THE 2-MINUTE PERFORMANCE UNDER BRIGHT, MEDIUM, AND DIM CONDITIONS

| Illumination | 30 sec. | 60 sec. | 90 sec. | 120 sec. |
|---|---|---|---|---|
| Bright | 18.2 | 18.3 | 18.9 | 18.5 |
| Medium | 17.2 | 16.2 | 16.9 | 15.6 |
| Dim | 13.3 | 15.1 | 15.4 | 14.7 |

and 120-second performance periods. The means are highest under the "Bright" condition and lowest under the "Dim" condition.

## DISCUSSION

The results show that the rate of reversal is affected by the degree of illumination, and that the brighter the illumination, the faster the rate of reversal. This is consistent with the prediction.

As far as absolute rate is concerned, the results are also consistent with the theory of satiation, yet they suggest that an additional explanation may be required to account for the acceleration of the rate of reversals.

Experimental evidence, supporting the hypothesis that the apparent reversal phenomenon is caused by satiation, hinges on studies (2, 3, 19) which showed an increasing rate of reversal or change as a function of time, and an approximately 100 per cent interocular transfer of this rate.

It must be pointed out, however, that in the above-mentioned studies, the accelerating rate was not found to be common to all subjects, and, for the study of transfer effects, only those subjects were used who showed a substantial acceleration.

An incidental but interesting finding is that only three subjects behaved according to the expectation that as illumination increased the rate-time curve would become more positively accelerated. As the degree of illumination increased, a smaller number of subjects showed an accelerating curve. Thus, under the Dim condition nine subjects showed this phenomenon; under the Medium condition, six; and under the Bright, only five. The rest showed either a fairly constant or a decelerating rate. In the over-all averages, the Dim condition was most favourable to a rapid acceleration of reversal rate. It accelerated rapidly up to the 60-second point then declined somewhat to a plateau. The average rate for the Medium illumination stayed fairly constant throughout the 120-second performance. For the Bright condition, there was again an accelerating curve (though much more gradual than in the Dim condition), which began above the Dim and Medium illumination points, reached an asymptote at the 100-second point, then began to taper off.

It appears, then, that degree of illumination plays a part not only in the absolute level of reversal rate but in the acceleration of this rate as well.


## Summary

This study was carried out to test the effects of illumination on the rate of reversal of an ambiguous figure. The hypothesis was that the rate of reversal would be an increasing function of illumination. This hypothesis was borne out by the results. Evidence was also given to indicate that there is an interaction between the degree of illumination and the acceleration of the reversal rate.


## REFERENCES


1. Bartley, S. H. The psychophysiology of vision. In S. S. Stevens (ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1951.
2. Brown, K. T. Factors affecting rate of apparent change in a dynamic ambiguous figure as a function of observation time. *WADC Technical Report* 53–482, Wright-Patterson Air Force Base, Dayton, Ohio, December, 1953.
3. Brown, K. T. Studies on rate of apparent change as a function of observation time using a new type of dynamic ambiguous figure. *WADC Technical Report* 54–139, Wright-Patterson Air Force Base, Dayton, Ohio, May, 1954.
4. Cochran, W. G., Autrey, K. M., & Cannon, C. Y. A double change-over design for dairy cattle feeding experiments. *J. Dairy Sci.*, 1941, **24**, 937–951.
5. Cochran, W. G., & Cox, G. M. *Experimental designs*. New York: Wiley, 1950.
6. Eysenck, H. J. Cyclothymia and schizothymia as a dimension of personality. *J. Pers.*, 1952, **20**, 345–384.
7. Eysenck, H. J., Holland, H., & Trouton, D. S. Drugs and personality. IV. The effects of stimulant and depressant drugs on the rate of fluctuation of a reversible perspective figure. *J. ment. Sci.*, 1957, **103**, 656–665.
8. Federer, Walter T. *Experimental design: Theory and application*. New York: Macmillan, 1955, pp. 438–443.
9. Grether, W. F., Cowles, J. T., & Jones, R. E. The effect of anoxia on visual illusions. *Psychological Research on Equipment Design, Report no. 19: Army Air Forces Aviation Psychology Program Research Reports*, 1947.
10. Guilford, J. P., & Braly, K. W. An experimental test of McDougall's theory of extroversion-introversion. *J. abnorm. soc. Psychol.*, 1931, **25**, 382–389.
11. Guilford, J. P., & Hunt, J. M. Some further experimental tests of McDougall's theory of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1932, **26**, 324–332.
12. Hunt, J. McV., & Guilford, J. P. Fluctuation of an ambiguous figure in dementia praecox and in manic depressive patients. *J. abnorm. soc. Psychol.*, 1933, **27**, 443–452.
13. Köhler, W. *Dynamics in psychology*. New York: Liveright, 1940.
14. Köhler, W., & Wallach, H. Figural after-effects: an investigation of visual processes. *Proc. Amer. Philos. Soc.*, 1944, **88**, 269–357.
15. McDougall, W. *Outline of abnormal psychology*. New York: Scribner's, 1926.
16. Mull, H. K., Armstrong, G., & Telfer, B. The effect of brightness upon reversible perspectives and retinal rivalry. *Amer. J. Psychol.*, 1956, **69**, 123–125.

17. Mull, H. K., Ord, N., & Locke, N. The effect of two brightness factors upon the rate of fluctuation of reversible perspectives. *Amer. J. Psychol.*, 1954, 67, 341–343.
18. Nemor, N. The alternation of an ambiguous figure in paretics and schizophrenics. *J. abnorm. soc. Psychol.*, 1953, 48, 445–446.
19. Wertheimer, M., & Jackson, C. W. Jr. Figural after-effects, "brain modifiability", and schizophrenia: a further study. *J. gen. Psychol.*, 1957, 57, 45–54.
20. Yacorzynski, G. K., & Davis, L. An experimental study of the functions of the frontal lobes in men. *Psychosom. Med.*, 1945, 7, 97–107.

# COLOUR JUDGMENT AS A FUNCTION OF STIMULUS CONDITIONS AND MEMORY COLOUR

ROBERT C. BOLLES
*University of Pennsylvania*
IRENE MACKINTOSH HULICKA AND BARBARA HANLY
*University of Oklahoma*

MUCH OF THE TIME perceptual cues to an object's colour are inconsistent. Thus, whenever the incident illumination is not "normal," there will be incongruence between the cues to spectral reflectance of an object (the distal stimulus) and the reflected light entering the eye (the proximal stimulus). Under these conditions subjects tend to show colour constancy: that is, perceived colour tends to be more highly correlated with the distal stimulus than with the proximal stimulus.

In discussing this phenomenon, Judd concludes that "color transformation and memory color work together to promote recognition of objects in unfamiliar illuminations and to prevent disorientation. In producing this effect they are aided greatly by retinal adaptation, and somewhat by simultaneous color contrast" (**10**, p. 859). There can be little doubt regarding the contribution to constancy of the sensory mechanisms mentioned by Judd, transformation, adaptation, and contrast; but the authors wish to question the contribution to colour constancy of the "central" factor, memory colour.

Hering defined memory colour as "the color in which we have most often seen a familiar object." He went on to make the observation that "provided we are not paying especial attention to an object's color, we will perceive it through the spectacles of our memory color" (**8**). This observation has been documented by Adams, who collected a number of instances where perceived colour was said to shift dramatically when an erroneously identified object was suddenly identified correctly (**1**). It should be noted that in many of these instances the incongruent perceived colour provided the basis for the correct identification, thus indicating the importance of Hering's proviso that attention should not be directed to the colour itself. Nonetheless, subsequent investigators have accorded an even more decisive role to memory colour by disregarding Hering's proviso.

These writers have asserted, like Judd, that perceived colour, as a

general rule, is a compromise between memory colour and the colour determined by sensory factors. A study by Duncker is sometimes cited as demonstrating such a compromise (5). Duncker found that a greyish paper in the shape of a leaf was matched as though it were greener than a spectrally identical paper in the shape of a donkey. According to many writers the leaf "looked" greener because what was seen was a compromise between the proximal stimulus, grey, and the distal stimulus, green, which was implied by the shape cues and mediated by memory colour.

But, the experimental evidence does not seem to require the conceptually complicated assumption that sensory and associative factors interact. An alternative interpretation of the available data is possible which is free of any interaction assumption. Under the physical conditions of Duncker's experiment it was impossible for an exact colour match to be made. He used green paper illuminated through a red filter for his standard stimuli and a green-orange-black-white colour wheel for his comparison stimuli. As a consequence of these different colour-producing operations, the stimuli to be matched probably differed in spectral composition, saturation, brightness, and surface texture. Other experimenters, too, in obtaining memory colour effects, have imposed impossible colour-matching conditions on their subjects (1, 2, 3, 4, 7). With no exact colour match possible, it seems reasonable to expect subjects with quite a range of possible mismatches to choose from, to pick one that "looks good," that is, one which tends to conform with memory colour.

The present experiments were undertaken to demonstrate that, if the subject is permitted to make a colour match, then he will do so regardless of possible memory colour influences. Memory colour effects (by which we mean mismatches in the direction of memory colour), we will argue, occur only when a psychophysical equation is impossible.

## Experiment I

In an earlier paper, one of the present writers found that subjects tended to follow the verbal labels, red, orange, and yellow, that had been previously associated with nonsense figures (all of the same colour) when matching the colours of the figures (2). These results were interpreted as evidence for the interaction of sensory and associative factors in perception. But the matching conditions that were used in that experiment involved grey paper figures against a blue-green background as standards and a red-yellow colour wheel as the comparison stimulus. The purpose of the present experiment was to determine if the basically impossible

task of obtaining a match under these conditions is a necessary requirement for obtaining the reported results.

## Method

The apparatus and procedure were the same as in the previous study except in one important respect. The one change was the introduction of 90° sectors of black and white into the colour wheel which had previously contained only red and yellow sectors. (Although it would have been preferable to use identical colour-producing conditions on both sides, the change that was made was appreciable, and enabled *E* to make satisfactory colour equations.) Since the method of the experiment has been previously described in detail, it will only be outlined here.

The *Ss* were 20 psychology undergraduates.[1] They were told that the experiment was on association, that they would be shown a number of figures paired with colour names, and that they would have to learn to say aloud the colour names when the figures were presented alone. After this association experience *Ss* made colour matches of the figures under the guise that that was part of another experiment which just happened to use the same figures.

Six different pairs of nonsense figures and colour names were presented 14 times each, in randomized order, to each *S*. Of the 6 figures, 3 were chosen to test the hypothesis that colour labelling alone might affect colour judgments and, therefore, these figures were always presented as neutral grey. One was paired with the word "yellow," another with the word "orange," and the third with the word "red." The three remaining figures were used to present colour as an attribute of form. One figure was cut from yellow paper, one from orange paper, and the third from red paper; each of these figures was paired with its appropriate colour name. Replicas of these coloured figures were made from grey paper. For half the presentations of each of these 3 figures the grey forms were used, and for the other half the figures were coloured. Figures and the associated colour names were displayed for 2-sec. intervals through windows in a large grey screen.

After the 84 pairings *Ss* were required to make a colour match for each figure. Colour comparisons were made successively rather than simultaneously (in an effort to maximize the role of non-sensory determinants of judgment). A figure was shown for 15 sec., removed, the colour wheel shown, *S* instructed *E* how to adjust the wheel for a better match, and the wheel was removed. This cycle was repeated until matches had been made for all 6 figures. The figures to be matched and the starting positions of the colour wheel were presented in a counterbalanced order.

## Results

The present results and a summary of the Baker and Mackintosh results are presented together in Table I. Table I gives the means and S.D.'s of the percentages of red used to match the six figures for each study.

The use of different colour wheel compositions precludes a meaningful comparison between the average colour match in the two situations, but other comparisons can be made. First, there is a marked reduction in

[1]This study was carried out by Hanly, under the direction of Hulicka, at the University of Oklahoma.

TABLE I

MEANS AND S.D.'S OF THE PERCENTAGES OF RED USED TO MATCH THE CONTRAST COLOUR OF FIGURES PREVIOUSLY ASSOCIATED WITH COLOURS

| | Colour as attribute | | | | | | Colour labelled | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Yellow | | Orange | | Red | | Yellow | | Orange | | Red | |
| | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D |
| Present results | 26.1 | 8.2 | 26.4 | 9.5 | 26.6 | 9.0 | 23.6 | 10.2 | 24.7 | 7.7 | 25.5 | 8.2 |
| Baker-Mackintosh results | 26.1 | 17.3 | 38.6 | 22.4 | 42.8 | 22.0 | 30.0 | 21.4 | 34.7 | 17.0 | 42.8 | 23.8 |

variability under the present conditions where an exact colour match could be more nearly approximated. The $F$ ratio of error variances of 5.31 is significant at the .01 level. This reduction in variability reflects the greater ease and confidence with which subjects can make their judgments.

But the critical finding here is that the influence of previous colour associations virtually disappears as the conditions for a colour equation are approached. Whereas Baker and Mackintosh found a highly significant over-all $F$ between figures, and four of their six relevant $t$ tests were significant, here the overall $F$ was less than unity, and no $t$ test was significant. These statistical tests yielded negative results in spite of the drastically reduced error variance. Again, it should be emphasized that the only difference between the two studies was the construction of the colour wheel which made a colour equation possible (or nearly so) in one case, and clearly impossible in the other.

## EXPERIMENT II

The purpose of the second experiment was, broadly speaking, to assess the generality of the stimulus conditions under which memory colour effects occur. Specifically, the aim was to test the following hypotheses:

(1) Memory colour effects depend upon colour-producing conditions that preclude an exact colour match. This is our principal hypothesis.

(2) Memory colour effects depend upon lack of "structure" in the test situation. Thus, the effects would tend to disappear if a stable reference or standard figure were always present in the matching situation.

(3) Memory colour effects are artifacts that depend upon purely physical properties of the stimulus figures such as area and perimeter-area ratio. The relevance of these physical determinants of perceived

colour in memory colour experiments has been recently emphasized by Fisher, Hull, and Holtz (6).

(4) Memory colour effects are artifacts that are due to the use of critically ambiguous colour-producing conditions. For example, the subject in matching an orange-coloured apple figure might perceive the orange hue as a yellowish red and attempt to match the red component.

The test situation was adapted from the one used by Duncker (5), and the experiment was factorially designed to test the four hypotheses just listed.

## Method

The subjects were 32 naïve college sophomores.[2] A two-chambered apparatus, following Duncker's dimensions, was constructed from white cardboard. Each side of the apparatus contained a 7½-watt red light bulb lighted by 110 volts, a 25-watt frosted white bulb lighted by a variable voltage, and a frame into which the stimulus figures could be inserted. *S* was seated about 18 in. in front of the stimulus figures where he could glance back and forth between them. *E* read the following instructions: "This is an experiment to see how well people can judge the colours of things. What you will do is look back and forth, from left to right, as often as you want, and compare the colours of the things on the two sides. While you are doing this, I will be slowly varying the colour of the thing on your left (right). When the colour of the two things gradually becomes the same, say stop. If we go too far in one direction, say so, and we will go slowly back. Here are the first two things to compare. I am changing the colour slowly now. Tell me when they are the same."

The illumination, and hence the colour of the figures, on one side of the apparatus was kept constant for a given *S* by operating the 7½-watt red bulb at 110 volts and the 25-watt white bulb at a fixed, but lower, voltage (for example, 60 volts). Figures on the other side of the apparatus were illuminated by operating the 7½-watt red bulb at 110 volts and the 25-watt white bulb at a voltage which might vary between 40 and 80 volts. The variable voltage was changed at the rate of approximately ½ volt per second; ascending and descending presentations and the starting voltages were randomized. Colour match errors were recorded in voltage differences between the two sides of the apparatus.

To test Hypothesis 1 Ss were randomly assigned to two groups, for one of which an exact colour match was possible because of the symmetry of the apparatus, while for the other group this symmetry was disturbed so that an exact match was impossible. The equality of stimulus colour on the two sides of the apparatus was altered by inserting sheets of yellow typing paper on one side of the centre partition of the apparatus, out of *S*'s view, so that the spectral composition of the reflected light reaching the figure and background on that side was yellower than on the other side. The stimulus figures used to test for memory colour effects were cut in the shapes of a leaf and a donkey, both copied from Duncker (5). In addition, three control figures were used, a rough-contoured nonsense figure, a smooth-contoured nonsense figure, and a square figure. The figures were each about 10 sq.

[2]This study was carried out by Bolles, at the University of California. The encouragement of Leo Postman is gratefully acknowledged.

cm. in area, and were cut from a pale green paper ("Color-Aid," G-Y-G). According to Hypothesis 1, under the match-impossible condition, the leaf figure should be matched as if it were greener than either the donkey figure or the control figures, that is, with a lower voltage on the white light bulb; while under the match-possible condition, no differences between figures should occur.

Each group was further divided into two subgroups that differed only in the manner in which the stimulus figures were presented. For one subgroup the square figure was always present and was paired with each of the other four figures four times in a counterbalanced order. For the other subgroup the square was not used; the leaf or the rough-contoured control figure was always paired with the donkey or the smooth-contoured control figure. Each of these four possible figure combinations was presented four times in a counterbalanced order. The purpose of these two subgroups was to provide a means of testing Hypothesis 2, that is, to determine if the presence of a reference stimulus (the square) in every set of figures to be matched would reduce the size or frequency of memory colour effects. For each subgroup the stimulus series was balanced for left-right and ascending-descending for each S, and was balanced across Ss for the side of fixed and variable colour.

Duncker's leaf and donkey figures differed not only in implied memory colour but also with respect to contour roughness (the leaf was serrated). Fisher *et al.* have shown that such physical features of stimuli can affect colour judgment (6). Hence, the use of rough-contoured and smooth-contoured control figures provided a basis for testing Hypothesis 3 which asserts that memory colour effects are artifacts depending upon physical stimulus properties.

Hypothesis 4 asserts that memory colour effects are artifacts that depend upon specific ambiguous colour conditions. To test this possibility, the colour of the standard stimulus was varied over a limited range across Ss. The reference voltage for the white bulb illuminating the standard stimulus, while fixed for a given S, was 50, 55, 60, or 65 volts for different Ss.

The overhead lights were off, so that the prevailing illumination was red. The stimulus figures themselves, within the range of illumination used, were quite desaturated, and varied from a dark reddish grey to a greenish grey. Through a reduction screen, the figures appeared achromatic throughout the range of illumination.

## Results

The colour-matching errors were subjected to an analysis of variance. Differences between figures were not significant, but the figure by subject interaction was significant, suggesting that only some subjects may have yielded memory colour effects. This finding, together with the marked individual differences in judgment variability, indicated the necessity of analysing the results of each subject individually.

Recall that the experimental design involved four judgments of each stimulus figure combination for each subject. Individual estimates of variability were made for each of these blocks of four judgments, and each of these variability estimates was divided by its respective mean judgment to give a standard score for each figure combination for each subject. None of the subjects judged the rough- and smooth-contoured

control figures significantly differently. This result enables us to reject Hypothesis 3, and to combine the data across the rough-smooth parameter to obtain more stable standard scores for the control figures. Those standard scores relevant to testing a leaf-donkey difference are presented in Table II.

TABLE II

INDIVIDUAL STANDARD SCORES FOR EACH COMBINATION OF STIMULI

| Subgroup | Figures | n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| Match-possible, no standard | L–C* | 4 | −1.51 | 2.10 | − .65 | .42 | −1.67 | − .11 | − .78 | 1.10 |
| | D–C | 4 | .62 | .14 | 1.68 | 1.16 | 1.19 | −1.31 | .27 | − .29 |
| | L–D | 4 | 1.20 | 1.47 | −1.62 | 2.44 | .68 | 1.89 | −1.89 | − .42 |
| Match-impossible, no standard | L–C | 4 | 1.13 | 2.88 | − .66 | − .68 | − .46 | .63 | 1.11 | − .63 |
| | D–C | 4 | − .65 | − .32 | − .62 | **6.00** | − .99 | − .30 | − .60 | −2.06 |
| | L–D | 4 | 1.61 | **7.00** | 2.06 | −1.77 | **7.70** | **5.46** | .44 | **5.20** |
| Match-possible, standard | C–S | 8 | − .48 | **−2.40** | −1.47 | .21 | 1.48 | .98 | − .72 | − .42 |
| | D–S | 4 | 1.26 | − .84 | .78 | 0.00 | .90 | .70 | 1.26 | .62 |
| | L–S | 4 | − .60 | −1.15 | − .77 | .88 | **3.89** | −1.09 | − .60 | .14 |
| Match-impossible, standard | C–S | 8 | 1.67 | −1.30 | − .71 | − .97 | .65 | **2.43** | .84 | .59 |
| | D–S | 4 | .92 | − .37 | .07 | .76 | .65 | 1.58 | 2.82 | .30 |
| | L–S | 4 | − .49 | **3.21** | −2.94 | .87 | −1.99 | 2.40 | 1.49 | − .39 |

*Figure abbreviations are: L = leaf, D = donkey, C = control nonsense figure, S = square. A "–" sign indicates that the right-hand member of a pair is judged to be the greener. Scores which are significant at the .05 level are given in bold face type.

In the "match-possible, no standard" subgroup there were no significant effects indicated for any figure combination by any subject. In the "match-possible, standard" subgroup subject no. 2 judged all the figures less green and the control figures significantly less green, than the square. Subject no. 5 showed a mean leaf-square match which significantly favours the leaf. But since all his figures tended to be greener than the square, and since the leaf greenness was not significantly greater than that of the other figures, it would seem gratuitous to attribute his results to a memory colour effect. In the "match-impossible, no standard" subgroups four subjects showed clear memory colour effects, while subject no. 4 gave anomalous results: for him the donkey seemed to be greener than the other figures. Finally, in the "match-impossible, standard" group one subject showed a significant memory colour effect, and subject no. 6 judged all the figures as though they were greener than the square.

To summarize these results with respect to Hypothesis 1, there was no evidence of memory colour effects under conditions where an exact colour match was possible, while five subjects out of sixteen showed effects under match-impossible conditions. Fisher's exact probability test (12), applied to the difference between proportions of positive

instances under match-possible and match-impossible conditions, indicates a probability of .022 that the observed difference is due solely to chance. The mean variability under the match-possible condition was significantly smaller than under the match-impossible condition, the *F* ratio being significant at the .01 level.

With respect to Hypothesis 2, the proportions of memory colour effects obtained under standard and no standard conditions were one and four subjects out of sixteen respectively. While the difference lies in the expected direction, it is not significant. Similarly, the standard condition yielded less variable judgments than the no standard condition, but not significantly less.

The variation in the fixed voltage at which the figures on one side of the apparatus were presented led to no significant source of variation in the initial analysis of variance, indicating that Hypothesis 4 was not confirmed in the present situation.

The results of the second experiment, then, like those of the first, clearly confirm Hypothesis 1 which asserts that memory colour effects depend upon stimulus conditions that preclude an exact colour match. The results also suggest, although not too convincingly, that even with impossible colour-matching conditions memory colour effects tend to disappear if a strong standard or reference stimulus is present (Hypothesis 2). This possibility deserves further investigation. Hypotheses 3 and 4, which assert that memory colour effects are artifacts, were not substantiated in the present situation.

In the post-experimental interviews, all subjects recalled the leaf and donkey, but there was no consistent identification of the nonsense figures. When asked what colours the figures "actually were," the subjects' replies varied from blue-green to black, but no one remembered them as pale green. Curiously, many subjects thought the different figures were actually different colours.

## Discussion

Previous discussions of memory colour have tended to interpret the memory colour effect as evidence that perception is a compromise between sensory factors and associative factors. Such an interpretation has been favoured by theorists who conceive of perception as an achievement or a product of the organism. In this view sensory and associative events interact to determine what is perceived. But this position is unintelligible to those who conceive of perception as a primary or antecedent process dependent only, in some psychophysical way, upon sensory events. This latter position must deny that perceived colour, for example, can be influenced by associative and experiential factors, admitting only that

judgments or responses about colours can be so affected (11). Many writers have noted that memory colour effects are enhanced by conditions that maximize associative or memory factors and minimize the stimulus support for colour judgment (1, 2, 3, 5). This kind of evidence would seem to lend further plausibility to the interaction assumption.

However, the evidence is not conclusive on this point because of the unfortunate confounding that has accidentally occurred of stimulus conditions that create ambiguity with stimulus conditions which preclude an exact colour match. The critical feature of the stimulus upon which the demonstration of memory colour effects depends may not be ambiguity *per se*, but rather whether or not a colour equation is possible at all. The results of the present experiments, like those of Bruner *et al.* (3), indicate that, even though the colour-matching task is a difficult one, involving desaturated contrast colours, the subject will make a colour match if it is possible to do so. If the experimental conditions prevent a match, then some subjects will choose a mismatch which tends to conform with the memory colour implied by the shape of the figure. The occurrence of a memory colour effect under these conditions does not have to be interpreted as evidence for a "distorted" perception of colour; it may only reflect the common desire of psychology undergraduates to do the "right" thing. Judgments that conform to internal standards of some sort are the only recourse possible in such a situation. In short, while there is ample evidence to show that memory colour can influence judgments about colour in the absence of the stimulus to be judged, there seems to be no necessity for supposing that memory colour influences phenomenal, immediately given, perception of colour.

It would seem then that the demonstrations of memory colour effects presently in the literature cannot properly be taken as evidence for colour constancy. The memory colour effect is a judgmental compromise which may be contrasted in several ways with the "regression toward the real object" typically found in constancy studies.

One important difference, for those who recognize a distinction between perceiving and responding, is that, so far, memory colour has only been shown to have an effect at the behavioural level, while constancy is manifest both phenomenally and behaviourally.

Second, the present results suggest that memory colour effects are restricted to match-impossible conditions. In constancy studies, on the other hand, perceptual compromises are the general rule over a wide range of stimulus conditions. It is true that the constancy ratio depends upon the richness of cues available (9), but the constancy compromise does not *require* test conditions that preclude zero constancy.

Third, compromises of perceptual constancy are a very general phe-

nomenon, shown by all subjects, while in the memory colour studies, where individual data are reported, only some subjects show effects.

Finally, compromises of perceptual constancy are generally of an appreciable magnitude relative to the range of possible judgments, while the memory colour effects reported here and in previous studies seem to be in the order of one or two j.n.d.'s (**1**, **2**, **3**). The size of the memory colour effect, according to the thesis of this paper, is limited by how nearly an exact colour match can be approximated. In view of these distinguishing features it does not seem reasonable to construe memory colour effects as instances of, or contributing to, colour constancy. Indeed, the two would seem to be unrelated.

## Summary

Two experiments are reported, one of which provides an important control for a study of Baker and Mackintosh involving colour judgments of nonsense figures previously associated with colour words. The conditions of the present experiment differed from the previous study in that they permitted rather than prevented an exact colour match. The second experiment, modelled after a study of Duncker's, involved two sets of conditions, which also differed in whether or not they permitted an exact colour match to be made.

The results indicate that the influence of associative factors or memory colour upon colour judgment is restricted to situations where an exact colour match is not possible. When a match can be made, S will make it regardless of associative influences. From this and other evidence we conclude that it is unlikely that memory colour contributes to colour constancy.

## REFERENCES

1. Adams, G. K. An experimental study of memory color and related phenomena. *Amer. J. Psychol.*, 1923, **34**, 359–407.
2. Baker, K. E., & Mackintosh, I. The influence of past associations upon attributive color judgments. *J. exp. Psychol.*, 1955, **49**, 281–286.
3. Bruner, J. S., Postman, L., & Rodrigues, J. Expectation and the perception of color. *Amer. J. Psychol.*, 1951, **64**, 216–227.
4. Cramer, T. Ueber die Beziehung des Zwischenmediums zu den Transformations- und Kontrasterscheinungen. *Z. Sinnephys.*, 1923, **54**, 215–242.
5. Duncker, K. The influence of past experience upon perceptual properties. *Amer. J. Psychol.*, 1939, **52**, 255–265.
6. Fisher, S. Carolyn, Hull, C., & Holtz, P. Past experience and perception: memory color. *Amer. J. Psychol.*, 1956, **69**, 546–560.
7. Harper, R. S. The perceptual modification of colored figures. *Amer. J. Psychol.*, 1953, **66**, 86–98.
8. Hering, E. *Grundzüge der Lehre vom Lichtsinn*, 1905 (quoted in Adams, **1**, p. 359.)

9. Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.
10. Judd, D. B. Basic correlates of the visual stimulus. In Stevens, S. S. (ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
11. Pratt, C. C. The role of past experience in visual perception. *J. Psychol.*, 1950, **30**, 85–107.
12. Siegel, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

# THE EFFECTS OF DRUGS ON A CONDITIONED AVOIDANCE RESPONSE[1]

GORDON A. McMURRAY AND L. B. JAQUES
*University of Saskatchewan*

THE SEARCH for behavioural methods to assess the action of drugs in animals has been largely a search for reliable response patterns that can differentiate between specific drug effects on learning or motivational-emotional factors, and the non-specific effects of motor disorder, sensory loss, sedation, and general malaise. One technique has been the conditioned avoidance and escape situation in which the animal performs some instrumental response which may anticipate a noxious stimulus (usually electric shock), thereby functioning as an avoidance response. If the response is made only to the noxious stimulus itself, then it is classified as an escape response. Since the avoidance and escape responses differ with respect to the eliciting stimulus, it is of interest to see if there is any differential effect of drugs upon them.

Other techniques include the conditioned emotional response, used extensively by Brady and Hunt (2), and the instrumental conditioning procedures introduced by Skinner (9), using different reinforcement schedules and a variety of discrimination problems. Sidman (8) has successfully used temporal patterning and discrimination for the further study of drug effects. These appear to be valuable methods for evoking response patterns which may show subtle changes under drug action and hence allow for the study of differences between drugs. The present study uses a variation of the conditioned avoidance-escape situation to observe the effects of chlorpromazine, reserpine, pentobarbital, secobarbital, atropine, scopolamine, dibenzyline, and ethyl alcohol.[2]

## METHOD

*Subjects and Apparatus*

The subjects were 93 male rats of Wistar stock obtained from a local dealer, ranging in weight from 140 to 210 gm. The apparatus consisted of a box made of ¾ in. plywood with over-all dimensions of 36¾ in. × 11¾ in. × 8 in. It was divided into two compartments of equal size by a sheet of aluminum 4 in. high set in metal slots on the side walls. One compartment, which was painted black, had a grid floor

[1]This study was supported by a grant from the Defence Research Board, Canada (Project DRB9350–02).

[2]The authors wish to thank Ciba Company Limited who supplied the reserpine, and Smith, Kline and French who supplied the dibenzyline used in this study.

of copper wire wound around a plastic insulator at ¾-in. intervals. The other compartment was painted white, and the floor was of the same smooth plywood as the walls. A power supply delivered 60 c/s current at 100 v., 1.5 — 2 m. amps., across the grid whenever a key was closed. The current resulting regularly elicited squealing, rapid foot withdrawals, and other escape movements from a rat in the shock compartment. The box was not covered, but a piece of plate glass was placed over the shock box immediately after the animal had been placed in this compartment.

*Procedure*

*Without drugs.* Each rat was given 5 min. to explore the two compartments. The first trial began immediately after, when the animal was placed in the black (shock) compartment and a timer started. At the end of 60 sec. shock was administered through the floor grid until the rat escaped by jumping over the low hurdle into the white compartment or by perching on the hurdle itself. The timer was stopped, measuring the time of escape, and a second timer was started to measure the 30-sec. interval preceding the next trial which began when the animal was picked up and replaced in the black compartment. If the rat avoided shock by leaving the shock box before the end of the 60-sec. period, the time of this avoidance response was noted and the next trial started 30 sec. later. In this procedure, escape (E) is the response to shock itself and avoidance (A) the response to stimuli anticipating shock. An animal failing to escape after an arbitrary time limit of 200 sec. was moved across the barrier by the experimenter, and the response recorded as escape failure (EF). The normal course of avoidance learning in this apparatus was studied in 10 undrugged rats given 35 to 70 trials each. Six of these trained animals were later used in testing chlorpromazine, but as a rule Ss used in the drug experiments were trained just prior to the administration of the drug.

*With drugs.* When drug effects were being tested the rats were trained by the above procedure, given a brief rest period, and then run for 15 to 35 additional trials to ensure the appearance of a normal pattern of avoidance responses. Each drug was administered by injection into a tail vein, and 15 to 30 min. later trials under drug conditions were begun. A second series of trials, started 90 min. after injection, was also given to 7 of the animals receiving reserpine. The following drugs were studied in this way: chlorpromazine (1 mg./kg.) in 12 rats; reserpine (1 mg./kg.) in 12 rats; pentobarbital (5 mg./kg.) in 5 rats; secobarbital (5 mg./kg.) in 6 rats; atropine (5 mg./kg.) in 6 rats; scopolamine (5 mg./kg.) in 6 rats; dibenzyline (5 mg./kg.) in 6 rats; ethyl alcohol (1 g./kg.) in 6 rats.

A further experiment was done in which rats were not trained prior to drug administration, but were first given the drug and the learning of the escape and avoidance response studied while the animal was under drug influence. Twelve rats were run 15 min. after chlorpromazine injection (1 mg./kg.); and 12 were run 20 and 90 min. after reserpine injection (1 mg./kg.).

## Results

Normal undrugged animals learned the avoidance response easily. Shocks were usually administered in the first three trials, and subsquently the rat avoided the shock by leaving the compartment before 60 seconds. Table I shows that, excluding the first three trials, a group of ten animals with no drug made 93 per cent successful avoidance responses (A) at

TABLE I

EFFECT OF DRUGS ON THE PERCENTAGE OF RESPONSES AS CONDITIONED AVOIDANCE (A), UNCONDITIONED ESCAPE (E), ESCAPE FAILURE (EF), AND MEDIAN TIME OF A IN CONDITIONED AVOIDANCE TRAINING

| Drug | Dose (mg./kg.) | *N* | Percentage before | | | Percentage after | | | Significance of change in percentage of A *P* | Median time of A | | Significance of change in median time of A *P* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | E | EF | A | E | EF | | Before | After | |
| No drug | — | 10 | 93 | 7 | 0 | — | — | — | — | 7.1 | — | — |
| Drug given after training | | | | | | | | | | | | |
| Chlorpromazine | 1 | 12 | 92 | 8 | 0 | 44 | 48 | 8 | <.01 | 5.0 | 14.4 | <.01 |
| Reserpine | 1 | 12 | 94 | 6 | 0 | 64 | 28 | 8 | <.01 | 2.6 | 6.6 | <.05 |
| Scopolamine | 5 | 6 | 96 | 4 | 0 | 30 | 38 | 32 | <.05 | 2.1 | 14.1 | — |
| Atropine | 5 | 6 | 96 | 4 | 0 | 94 | 6 | 0 | | 2.0 | 2.0 | |
| Dibenzyline | 5 | 6 | 98 | 2 | 0 | 99 | 1 | 0 | >.05 | 1.5 | 1.7 | >.05 |
| Ethyl alcohol | 1000 | 6 | 99 | 1 | 0 | 98 | 2 | 0 | | 1.5 | 1.8 | |
| Pentobarbital | 5 | 5 | 83 | 17 | 0 | 80 | 19 | 1 | | 2.3 | 5.1 | |
| Secobarbital | 5 | 6 | 93 | 7 | 0 | 97 | 3 | 0 | >.05 | 6.6 | 2.6 | >.05 |
| Drugs given before training | | | | | | | | | | | | |
| Chlorpromazine | 1 | 12 | — | — | — | 20 | 54 | 26 | — | — | 39.9 | — |
| Reserpine | 1 | 12 | — | — | — | 66 | 21 | 13 | — | — | 10.9 | — |

NOTE: All tests of significance were made by the Wilcoxon Matched-pairs Signed-ranks Test.

a median time of 7.1 seconds. This was over a block of 32 trials each. These results were repeated in the preliminary training before drug administration when characteristically about 94 per cent avoidance responses were obtained at a median time always under ten seconds and usually under three.

Chlorpromazine very significantly reduced the percentage of avoidance response (A) and increased the median time of those that were made. Although the proportion of successful avoidance responses was much reduced, the rats still responded effectively to shock itself. This is shown in Table II. This preservation of an escape response without producing a large percentage of escape failures indicates that the effect on the avoidance response was not due to gross motor or sensory disturbances.

TABLE II

EFFECT OF CHLORPROMAZINE ON MEDIUM ESCAPE TIMES

| Rat no. | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Median escape time | | | | | | | | | | | | |
| Undrugged | * | .6 | 7.0 | 3.4 | 3.2 | 3.6 | 2.5 | 2.8 | .7 | 2.0 | 1.8 | 2.4 |
| Drugged | 6.5 | 19.5 | 140.0 | 15.2 | 1.8 | 8.1 | 1.3 | 2.8 | 1.0 | 1.3 | 1.3 | 1.2 |

NOTE: The Wilcoxon Matched-pairs Signed-ranks Test showed non-significant difference between median escape times drugged *vs.* undrugged, $p > .50$.
*No value in records for this rat.

Reserpine also depressed the avoidance response without markedly increasing the percentage of escape failure (Table I). Although a very significant reduction in percentage of avoidance occurred, the action of reserpine appeared more variable than that of chlorpromazine. Four of the twelve rats were affected only slightly, and the median time of avoidance after drug was not markedly increased. Seven of the animals receiving reserpine were run again 90 minutes after drug injection and showed 63 per cent avoidance responses, very nearly the same as shown 15 minutes after the drug (Table I).

When chlorpromazine was given before training, it was very difficult to establish a conditioned avoidance, and even the executing of a prompt escape pattern was markedly retarded. With reserpine, some avoidance learning occurred, but the efficiency was very significantly below that of the undrugged animal. Apparently, when these drugs are given before the animal has learned to avoid the shock, the acquisition of both avoidance and escape responses is retarded.

Table I shows that scopolamine also significantly reduced the percentage of avoidance responses. But the large percentage of escape failures indicates that, unlike the results with chlorpromazine and reser-

pine, this result may be due to general sedation. All other drugs tested failed to produce a significant decrease in percentage of avoidance responses. This result appears particularly noteworthy in the case of barbiturates. These drugs as administered produced marked ataxia and sedation, but as long as the animals were able to move they quite successfully avoided the shock.

## Discussion

The results show that chlorpromazine and reserpine significantly reduced the number of avoidance responses while leaving the animal capable of behavioural arousal and of escaping the shock. This action of chlorpromazine and reserpine is markedly different from that of pentobarbital and secobarbital which leave the avoidance response relatively unaffected as long as the rat is capable of response. It is also different from the action of the other drugs which, at the dosage tested, had little effect on either avoidance or escape. Scopolamine also reduces the percentage of avoidance responses, but the marked rise in escape failure (EF) responses indicates that this may be due to general sedation rather than a specific effect on the avoidance response.

This "tranquillizer" effect on avoidance responses has been noted by other investigators. Weiskrantz and Wilson (**10**) showed in monkeys that reserpine blocked a lever-pressing response which delayed the occurrence of shock for ten seconds. The drug had the effect of preventing the anticipatory response to shock although frequent reactions to shock itself were noted. Cook and Weidley (**4**) used a technique in which rats could avoid shock by climbing a pole at the sound of a buzzer, previously associated with shock. Chlorpromazine and reserpine selectively blocked this avoidance response while the escape response (pole climbing after shock) remained relatively intact. Brady (**1**) also reports that the reserpine can eliminate the usual effects of a conditioned emotional response on cumulative response curves for lever-pressing in rats.

These results are interpretable in terms of the hypothesis that (1) fear responses are conditioned to stimuli contiguous with shock and (2) this conditioned fear supplies the stimuli which evoke an avoidance response reinforced through reduction of the fear drive. It would be expected then that agents which either reduce the strength of the conditioned fear drive or, as Heistad (**5**) has suggested, markedly change and distort the fear response-produced stimuli would result in loss of conditioned avoidance responses. It is possible that chlorpromazine and reserpine could have such effects on fear.

It should be noted, however, that this blocking of conditioned responses is not restricted to responses conditioned to aversive stimuli.

Weiskrantz (10) found that reserpine practically eliminated lever-pressing motivated by food reward. Olds (7) with a new drug-testing technique showed that chlorpromazine and reserpine reduced lever-pressing which delivered electrical stimulation to hypothalamic and amygdaloid areas. In the undrugged rat such stimulation in these areas acts as a reward, maintaining lever-pressing rates as high as 5,000 per hour.

Hunt (6) cast some further doubt on this "fear" hypothesis when he exposed rats previously injected subcutaneously with chlorpromazine (10 mg./kg.) to a clicking noise terminated by two inescapable shocks. Though animals with this dosage could show little response in this situation, it was found 48 to 50 hours after the last dose that the rats responded with typical freezing and immobility to the clicking noise. This indicates that associative learning could occur even under heavy chlorpromazine dosage although the fear response so learned is weaker than normal (6). It is still possible, when actually under the influence of the drug, that the fear response might be so radically affected that it cannot exert its motivating function in the conditioned avoidance response.

It would seem then that, although the "tranquillizers" may have a special effect on fear-motivated behaviour, their action is not confined to this alone. Brodie (3) has suggested that reserpine through the release of serotonin in the brain activates parasympathetic centres in the hypothalamus, allowing them to predominate over sympathetic-like effects. Chlorpromazine produces a similar result probably by blocking chemical mediators activating sympathetic centres. Thus, through different mechanisms, it is likely that these drugs markedly change hypothalamic functioning, tending to reduce highly affective-motivational components. It is this action which may markedly depress the performance of responses which are evoked by stimuli associated with motivational conditions, such as fear or hunger.

## Summary

Rats were trained on a conditioned avoidance response of the barrier-crossing type and then tested for the effects of chlorpromazine, reserpine, scopolamine, pentobarbital, secobarbital, atropine, dibenzyline, and ethyl alcohol. The learning of this response by previously untrained rats while under the effects of chlorpromazine and reserpine was also studied.

Chlorpromazine and reserpine significantly reduced the percentage of successful avoidance responses although the escape was still made effectively by most animals. Scopalamine also reduced the percentage of avoidance responses, but, unlike chlorpromazine and reserpine, markedly increased the percentage of escape failure. Pentobarbital and secobarbital had little effect on avoidance behaviour as long as the animal was capable of responding. Atropine, dibenzyline, and ethyl alcohol in the dosages used had little effect on either the avoidance or the escape responses.

## REFERENCES

1. Brady, J. V. A review of comparative behavioral pharmacology. *Ann. N.Y. Acad. Sci.*, 1957, **66**(3), 719–732.
2. Brady, J. V., & Hunt, H. F. An experimental approach to the analysis of emotional behavior. *J. Psychol.*, 1955, **40**, 313–324.
3. Brodie, B. B., & Shore, P. A. A concept for a role of serotonin and norepinephrine as chemical mediators in the brain. *Ann. N.Y. Acad. Sci.*, 1957, **66**(3), 631–642.
4. Cook, L., & Weidley, E. Behavioral effects of some psychopharmacological agents. *Ann. N.Y. Acad. Sci.*, 1957, **66**(3), 740–752.
5. Heistad, G. T. An effect of electroconvulsive shock on a conditional avoidance response. *J. comp. physiol. Psychol.*, 1955, **48**, 482–487.
6. Hunt, H. F. Some effects of drugs on classical (Type S) conditioning. *Ann. N.Y. Acad. Sci.*, 1956, **65**(4), 258–267.
7. Olds, J., Killam, K. F., & Bach-Y-Rita, P. Self-stimulation of the brain used as a screening method for tranquilizing drugs. *Science*, 1956, **124**, 265–266.
8. Sidman, M. Technique for assessing the effects of drugs on timing behavior. *Science*, 1955, **122**, 925.
9. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
10. Weiskrantz, L., & Wilson, W. A. The effects of reserpine on emotional behavior of normal and brain-operated monkeys. *Ann. N.Y. Acad. Sci.*, 1955, **61**(1), 36–55.

# THE SOCIAL DESIRABILITY OF TRAIT DESCRIPTIVE TERMS: ORDER AND CONTEXT EFFECTS[1]

EMORY L. COWEN
*University of Rochester*
AND
ALFRED STILLER
*University of Buffalo*

WITHIN THE AREA of personality assessment and research we are becoming increasingly concerned with formal and stylistic components of response, as opposed to content elements (4). In particular, there has been rapid growth in the study of the contaminating effects of social desirability (S-D) stereotypes in many widely used assessment procedures (2). Most S-D researches have been aimed at practical and applied issues, for example, the influence of this variable on a given test, or the control or reduction of its artifactual effects.

Recently, S-D norms for a large pool of trait descriptive terms have been presented (1). At that time, the need for information about the "basic variables—organismic, situational and personality, of which S-D stereotypes may be a function"—was stressed, and one parameter, sex differences, was studied. Females were found to rate positive adjectives more positively and negative adjectives more negatively than did males. Males, however, were significantly more variable in their S-D perceptions.

The present research focuses on the variables of order and context as factors influencing S-D perceptions. In the original normative study the context used was a large pool of trait descriptive terms, ranging from highly positive to highly negative adjectives and including intermediate ones. Here, we have extracted limited sub-pools of homogeneously positive and homogeneously neutral terms for S-D rating by a new group of comparable subjects. In accordance with the principle of the "adaptation-level" (3), implying variations in general frame of reference as a function of over-all context, we have hypothesized that positive adjectives, presented in a small homogeneous group as the initial rating task, will tend to be rated lower than they had been rated in the larger context of the normative study. Neutral adjectives, when given as the initial rating task, should have means comparable to those obtained in the normative study.

[1]Portions of this paper were presented at the annual meetings of the Eastern Psychological Association, Atlantic City, N.J., April, 1959.

It was planned that half the group would rate the positive adjectives first and then the neutrals, while the other half would do exactly the opposite. On the basis of an anchoring (5) or frame of reference position (7) we expected that positive adjectives given first should be rated lower than the same adjectives given second, while neutral adjectives given first would have higher ratings than neutrals given second.

## Method

On the basis of the original normative study (1) mean S-D values were available for a pool of 209 trait descriptive adjectives. From these data 2 smaller samples of items ($N = 30$ each) were constituted. The "positive" pool consisted of adjectives ranging in mean S-D from 6.07 to 6.77 ($X = 6.35$) on a 7-point scale (with 7 representing maximal S-D). The "neutral" pool was made up of adjectives ranging from 3.05 to 4.89 ($\bar{X} = 4.03$) in the normative study. Only adjectives which did not have significant sex differences in the normative study were used.

Subjects consisted of 57 undergraduates (19 male and 38 female), in advanced psychology courses at the University of Rochester. The task was group-administered to 3 groups of about 20 Ss each. Thirty-two Ss rated the positive adjectives first followed by the neutral ones, while the remaining 25 Ss rated the neutrals first and then the positives.

## Results and Discussion

### *Sex Differences*

In order to determine the feasibility of pooling the data for the two sexes, $t$ ratios were computed testing the significance of the difference of S-D means across orders, for males and females. This was done individually for each of the 30 positive and 30 neutral adjectives.

Two of the 30 adjectives in each pool yielded significant sex differences, with females scoring higher in all four instances. On this basis the ratings of the two sexes were combined for the later analyses of order and context effects.

However, disregarding these significant $t$ ratios for sex differences on individual adjectives, and looking at the directional pattern, we find that females had higher means on 27 of the 30 positive adjectives, with one even, and males higher on two. The probability of occurrence of such a distribution by chance alone, as tested by $\chi^2$ is less than .001. Both the $t$ ratio and directional analyses are entirely consistent with the comparable analyses reported in the original normative study (1).

### *Variations in Mean S-D Levels*

For each adjective an individual item mean S-D value (combining the two sexes) was determined. This was done separately for the two orders of administration. Table I presents the over-all mean S-D and the S-D range for both orders for positive and neutral adjectives, along with comparable data from the normative study (1).

TABLE I

S-D VALUES FOR POSITIVE AND NEUTRAL ADJECTIVES BASED ON ORDER 1, ORDER 2, AND NORMATIVE DATA

| | Order 1 | | | Normative | | | Order 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Mean | Range | *N* | Mean | Range | *N* | Mean | Range |
| Positive | 32 | 6.26 | 5.69–6.69 | 67 | 6.35 | 6.07–6.77 | 25 | 6.63 | 6.12–6.92 |
| Neutral | 25 | 4.12 | 2.76–5.16 | 67 | 4.03 | 3.05–4.89 | 32 | 3.76 | 2.53–4.81 |

By comparing mean S-D values for the 30 positive and neutral adjectives when they appear first with the mean S-D values for the same adjectives from the normative study, a test of the context effect is provided. The rating required in the present study is exactly the same as in the normative study except that here the subject is asked to rate 30 homogeneous adjectives rather than the same 30 in a context of 209 dispersed ones.

The primary order effect test comes from the direct comparison of the mean S-D values for the first and second orders, for both positive and neutral adjectives. Here the critical independent variable is the effect of the preceding task upon the second set of ratings. Mixed order and context effects would be considered to be present if significant differences were noted between the mean normative S-D values and those obtained for the same adjectives in the second order for either sub-pool. Table II presents all combinations of *t* ratios testing the significance of the above differences.

*Context Effects*

Positive adjectives when given first have slightly lower over-all mean values than the same adjectives in the normative study, while the neutrals

TABLE II

*T* RATIOS COMPARING MEAN S-D VALUES FOR ORDER 1, ORDER 2, AND NORMATIVE STUDY

| | *Dm* | *t* | *p* |
|---|---|---|---|
| *Positive adjectives* | | | |
| Order 1 *vs.* Normative* | .09 | 1.50 | N.S. |
| Order 1 *vs.* Order 2* | .37 | 5.29 | $<.001$ |
| Order 2* *vs.* Normative | .28 | 4.67 | $<.001$ |
| *Neutral adjectives* | | | |
| Order 1* *vs.* Normative | .09 | 0.50 | N.S. |
| Order 1* *vs.* Order 2 | .36 | 2.00 | $<.05$ |
| Order 2 *vs.* Normative* | .27 | 1.69 | N.S. |

*Higher mean

given first are slightly higher. Neither of these differences is statistically significant. A directional analysis based on individual adjective means, however, indicates that 20 of the positive adjectives have higher S-D means in the normative study than they did in Order 1. The reverse was true for nine adjectives, while one was the same both times. This distribution yields a $\chi^2$ value of 4.04 ($p < .05$). For the neutral adjectives, 22 of the 30 had higher Order 1 means as compared to their means in the normative study. The resultant $\chi^2$ is 6.54 ($p < .02$).

On the basis of these non-parametric analyses we observe that positive adjectives are rated lower when the context is a small homogeneous group than when they are interspersed among a sizable number of adjectives including neutral and negative ones. In the latter situation the probability of relatively more favourable perception is increased by the contrast value afforded by the presence of negative terms in the rating series. Neutral adjectives, however, were rated higher in Order 1 than in the normative study. This finding may be an artifact of two other circumstances: (1) means for the neutral adjectives are more dispersed in Order 1 than in the normative study (see Table I); (2) the 30 neutral adjectives include 18 slightly positive and 12 slightly negative items (rather than the chance split of 15–15). Thus the increased variability would act to raise more means than it would lower.[2]

*Order Effects*

For the positive adjectives the over-all mean S-D values in Order 2 are significantly higher than for Order 1. The reverse is true for the neutral adjectives where Order 1 means are significantly higher. Both of these parametric analyses are strongly supported by the results of the directional analyses. For all 30 positive adjectives, Order 2 means are higher than the comparable Order 1 values ($p < .001$), while for neutral adjectives 26 of the 30 Order 1 means are higher ($p < .001$). On the basis of these findings it seems clear that the initial rating task becomes an anchor point or frame of reference for the second. Fundamentally positive adjectives become even more socially desirable when they are rated after neutral ones. These data are in accord with those of Shapiro and Tagiuri (6) who note a context effect in connection with the definiteness of inferences made about traits. These authors state "the context effect we have observed is essentially one of contrast."

Neutral adjectives, when rated after positive ones, are seen as more negative than when they are rated first. Quite probably this is a function

[2]Fourteen of the eighteen slightly positive adjectives did actually have higher means in Order 1 than they did in the normative study ($p < .02$).

of the preceding positive anchoring experience and the relatively large contrast effect.

*Mixed Effects*

Positive adjectives given second had significantly higher S-D means than did the same adjectives in the normative study. Neutral adjectives had lower means in the second order as compared to the normative study: this difference approached but failed to reach statistical significance. Directionally, 29 of the 30 positive adjectives had higher means on the second administration than they had in the normative study ($p < .001$), while 23 of the neutral adjectives had lower means in the second order than they had in the normative study ($p < .001$). These findings, although statistically significant, add little to what we already know from the previous separate analyses of order and context effects.

*Intercorrelations*

The mean S-D values for the 30 positive and 30 neutral adjectives were intercorrelated for the three sets of data, providing a total of six correlations. The computed *r*s were unexpectedly high, ranging from .80 to .82, and suggested a relative constancy in S-D perceptions within these types of rating situations. This is so notwithstanding the constricted range of S-D means, due to the use of homogeneous trait pools and the large absolute shifts produced by order and context changes.

*Changes in Variability*

Although analyses comparable to those reported for mean S-D changes were also made for changes in variability, the latter will be summarized only briefly. We note first some tendency towards a greater dispersion of individual adjective means (see Table I) for Orders 1 and 2 in comparison with the normative study. When we examine more specifically the variances of individual adjectives for the three conditions, the most striking finding is that for 25 of the 30 neutral adjectives there is greater variability for Order 1 compared with the normative data ($\chi^2 = 13.3$, $p < .01$). Also, for neutral adjectives in 21 of the 30 instances, Order I produces greater variability than Order 2 ($\chi^2 = 4.8$, $p < .05$). The first of these findings suggests that when subjects are confronted with a homogeneous group of neutral adjectives, and lack the context of accompanying positive and negative adjectives which were available in the normative study, they tend to use a greater variety of ratings and draw more upon fairly positive and fairly negative ratings. The latter were not used for the same adjectives in the normative study because there were relatively more suitable adjectives present to receive those ratings.

The second finding illustrates the same principle to a somewhat lesser extent; in this case, however, the antecedent rating of a homogeneous group of positive adjectives may diminish somewhat the inclination to use fairly positive ratings, so that the effect is slightly less clear-cut.

*Implications for S-D Research*

The data from the present study suggest quite strongly that modifications of the absolute values of S-D ratings are produced by variations in order and context. Although the practical effects of these findings are restricted somewhat by the demonstration of considerable constancy in the relative ordering of the concepts across the several conditions, the data nevertheless indicate that systematic alterations of order and context are sufficiently important to alter appreciably the baseline normative attributes of trait-descriptive terms. To the extent that delimited pools are needed in subsequent substantive research, S-D values for trait-descriptive terms should be established in contexts comparable to the final test conditions under which they are to be used.

## Summary

Fifty-seven undergraduates rated two homogeneous sets of 30 trait-descriptive adjectives on a seven-point scale of social desirability. One set consisted of highly positive terms while the other was made up of neutral adjectives. These judgments were based on normative data collected previously. Approximately half the subjects rated the positive adjectives first followed by the neutrals, while the remainder did the opposite.

A context effect was demonstrated in that positive adjectives rated first were seen as less socially desirable, while neutral adjectives rated first were seen as more socially desirable, than were the same adjectives in the normative setting. Marked order effects were present, such that positive adjectives rated after neutrals were seen as consistently more desirable than the same adjectives rated before neutrals. Similarly, neutral adjectives given first were rated consistently higher than the same neutral adjectives following positive ones. In both orders neutral adjectives were rated with greater variability than in the normative study.

It seems probable that the variables of order and context systematically affect the absolute but not the relative perceptions of social desirability.

## REFERENCES

1. Cowen, E. L. The social desirability of trait descriptive terms: Preliminary norms and sex differences. *J. soc. Psychol.*, 1959. (In press)
2. Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden Press, 1957.
3. Helson, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

4. JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.
5. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, **37**, 1–42. Whole no. 261.
6. SHAPIRO, D., & TAGIURI, R. Some effects of response context on trait inference. *J. Personal.*, 1958, **26**, 42–50.
7. SHERIF, M. *The psychology of social norms.* New York: Harpers, 1936.

# COGNITIVE EFFECTS OF PERCEPTUAL ISOLATION[1]

T. H. SCOTT,[2] W. H. BEXTON,[3] W. HERON AND B. K. DOANE[4]

*McGill University*

IN EARLIER PRELIMINARY REPORTS (1, 3, 4) we have described some of the effects of exposing human subjects to a relatively unchanging sensory environment. The present paper reports a more systematic investigation of the effects of perceptual isolation on certain tests of cognitive function given during the experimental period and shortly after the subject had returned to a normal environment.

## METHOD

The subject was paid to lie 24 hrs. a day on a comfortable bed in a lighted, semi-soundproof cubicle, 8 ft. × 4 ft. × 6 ft., which had an observation window. Throughout the experiment he wore translucent goggles which admitted diffuse light but prevented pattern vision. Except when eating or at the toilet, he also wore cotton gloves and cardboard cuffs which extended from below the elbows to beyond the fingertips, so as to limit tactual perception. A U-shaped foam-rubber pillow, the walls of the cubicle, the masking noise of the thermostatically regulated air-conditioner and other equipment severely limited auditory perception. A two-way speaker system allowed communication between *S* and *E*.

The subjects were asked to stay as long as they could (usually 3 to 4 days), and during this period were prevented as far as possible from finding out what time it was. An experimenter was always in attendance, and Ss were told that if they needed anything they had only to call for it. They were fed and went to the toilet on request. These breaks occupied, on the average, 2 to 3 hrs. a day.

### *Tests*

We used two batteries of tests, one given during the isolation period, and one afterwards. We also tried to measure S's susceptibility to propaganda, as another means of assessing the cognitive effects of isolation.

The first set of tests ("cubicle battery") had two parts: A, five types of problems taken from various intelligence tests; and B, associative learning and digit span from the Wechsler memory battery, as well as an analogies test. The tests in Part A were given before, during, and after the isolation period, those in Part B were given shortly before the subject left isolation. The tests in the cubicle battery were given orally, and the subject did them in his head.

The types of problem in Part A of the battery were (1) multiplication of two and three digit numbers; (2) arithmetic "catch" problems, such as "how many

---

[1]D. R. B. Project no. D 77–94–85–01.

[2]Now at the University of Auckland, New Zealand.

[3]Now at Waterloo College.

[4]Now at the Faculty of Medicine, Dalhousie University.

times greater is twice 2¼ than half of 2½?"; (3) number series completion; (4) word-making, in which S had to make as many words as he could using the letters of a given word, and following certain rules (he could not make the same word twice, use proper nouns, and so on); and (5) anagrams, in which he had to make a word from a group of jumbled letters.

We took great care to make sure that S was awake some time before he was tested. All problems were read to him twice, and, except in the arithmetic "catch" problems and number series completion, he had to repeat the problem aloud to prove that he had heard it correctly. He was told that if he forgot the problem it would be repeated on request.

In the first three types of problem in Part A, S worked until he had the correct answer, or until 8 min. were up. In word-making he worked for 5 min. In the anagrams, if the correct word was not formed in 4 min., its first letter was supplied, subsequent letters being given at 3-min. intervals when necessary. Three scores were obtained from each test: time spent on the problem, number of wrong answers, and the number of requests for the problem to be repeated.

The second ("post-cubicle") battery contained the following tests: Kohs' blocks, Wechsler digit-symbol, Thurstone-Gottschaldt figures, transcribing a passage of unfamiliar technical material, the McGill Picture Anomaly test, and an unpublished block-design test (Delta Blocks) described by Hebb (2).

*Propaganda*

The propaganda material dealt with psychical phenomena. It consisted of a 90-min. talk read in a deliberately boring monotone, and was recorded on discs. Although there were general arguments in favour of believing in all types of psychical phenomena, four main topics were concentrated upon: telepathy, clairvoyance, ghosts, and poltergeists. In addition, the general importance of psychical research was discussed.

To measure attitudes towards psychical phenomena, we used a questionnaire consisting of a series of Bogardus-type scales. The questionnaire was divided into five sections, dealing with attitudes towards telepathy, clairvoyance, ghosts, poltergeists, and psychical research. There was a total of fourteen scales in the questionnaire, three in each of the first four sections and two in the fifth. Each scale consisted of five statements indicating different degrees of belief in the relevant phenomenon.

For scoring purposes, a three-point interval was set between successive statements on the scale. S had to indicate the statement with which he agreed most. If he was unable to decide between two statements, the average weight of the statements concerned was ascertained. A score of zero indicated that S did not believe in the particular psychical phenomenon at all, while a score of twelve indicated that he believed in it firmly. Since each of the first four sections of the questionnaire consisted of three scales, complete belief in some subject (telepathy, for example) would be indicated by a score of 36, complete disbelief by a score of zero. A score of 18 would, of course, indicate uncertainty.

Each section of the questionnaire also had a secondary series of scales. These were concerned with the amount of interest which S felt in the topic, and how important he felt the topic was.

*Subjects*

The subjects were English-speaking male college students. The mean age of the experimental group was 22 years, 1 month (range 19–30), of the control group 22 years, 2 months (range 19–32). Each group had the same proportion of Arts and Science students. Of the 29 experimental Ss, only 18 stayed in isolation long

enough to complete the testing schedule for Part A of the cubicle battery. Twenty-seven control Ss were tested on both parts of this battery.

There was some variation in the number of Ss on each test in the post-cubicle battery, as shown in Table II, since some tests were added to this battery later.

Of the 29 experimental Ss 24 were in isolation long enough to be given the propaganda. Data from 8 additional control Ss are available on the propaganda questionnaire. Though these data did not alter the statistical probabilities of the results, they are included, so that the control group for this test totals 35.

*Procedure*

The testing schedule of the experimental Ss is shown in Table I. The control Ss were given the same tests at the same intervals; they were not, of course, placed in isolation, but came to the laboratory at the appropriate times and were tested in a quiet room.

The subjects were given the questionnaire twice before they were placed in isolation. This was done because preliminary experiments showed that the reliability of the questionnaire was much greater between the second and third administrations than between the first and second.

The propaganda was given as follows: after S had been in isolation for approximately 18 hrs. he was told that there was a series of records that he could listen to if he wished. If he asked for a record, only one would be played at that time, but he could ask for another whenever he wished. He was not told anything about the content of the records. The nine records were played through in series as S

TABLE I

SEQUENCE OF TESTS GIVEN BEFORE, DURING, AND AFTER ISOLATION

| | Pre-isolation | | |
|---|---|---|---|
| 1-2 weeks | 48 hours | 24 hours | Immediately pre-isolation |
| Interview | First form of post-cubicle battery | Practice on cubicle battery A | Cubicle battery A |
| First attitude questionnaire | | Test on cubicle battery B | Second attitude questionnaire |
| | Isolation | | |
| 12 hours | 24 hours | 48 hours | 1 hour pre-emergence |
| Cubicle battery A | Cubicle battery A | Cubicle battery A | Cubicle battery B |
| | Records | Records | |
| | Post-isolation | | |
| Immediately | 3 hours | 3 days | 4 days |
| Post-cubicle battery | Third attitude questionnaire | Cubicle battery A | Cubicle battery A |
| | Interview | Interview | |

requested them. When they were finished he was told, "Those are all the records. You may have any of them you want played through again." On further requests, if no specific record was mentioned, the series was given in the original order, one record for each request.

The control subject was treated in a similar way. He was told about the records at approximately the same time in the testing schedule as the experimental Ss, and was made to hear the records at least once. He did not have to do this all at one sitting, but could hear any of the records as often as he liked during his subsequent visists. He was put in a room with the records and a machine for playing them, and listened to them through ear-phones. The attitude questionnaire was not given for the third and final time until the subject had definitely stated that he did not want to hear any of the records again.

## Results

Table II shows the results of the post-cubicle battery of tests. The $p$-values are for the difference between the mean scores of the experimental and control subjects immediately after isolation, relative to their scores before isolation. The experimental subjects were inferior to the controls on six of the seven tests. Only in the case of mirror drawing was there no difference between the groups.

The five graphs in Figure 1 show the mean time scores (except in the word-making test, where number of correct words is given) for the experimental and control groups on the five types of problem in cubicle battery Part A for each of the six test periods. A statistical analysis based on a comparison of the mean difference scores (between the pre-isolation and subsequent test periods) of the two groups reveals the following differences: on the word-making test the experimental group was significantly poorer on all three test periods during isolation ($p < .02$), and on the number series test poorer on the first cubicle test period

TABLE II

Mean Scores of the Experimental and Control Groups on the Post-Cubicle Battery: $p$-values based on Difference Scores

| Test | Score basis | Experiment group | | | Control group | | | |
|---|---|---|---|---|---|---|---|---|
| | | $N$ | Before | After | $N$ | 1st test | 2nd test | $p$ |
| Kohs' Blocks | Total time (sec.)* | 20 | 1088 | 931 | 25 | 1095 | 762 | .01 |
| Digit symbol | Number correct | 19 | 52.9 | 68.2 | 24 | 52.0 | 74.5 | .01 |
| Thurstone-Gottschaldt | Number correct | 12 | 5.5 | 5.4 | 18 | 5.2 | 8.1 | .01 |
| Copy passage | Time (sec.) | 18 | 594 | 640 | 25 | 634 | 639 | .05 |
| Delta Blocks | Number correct | 12 | 9.4 | 13.2 | 19 | 11.4 | 19.9 | .01 |
| Picture anomaly | Number errors | 15 | 3.0 | 5.9 | 23 | 4.0 | 4.9 | .01 |
| Mirror drawing | Time | 12 | 219 | 108 | 19 | 223 | 103 | .10 |

*The number of Kohs' items on which the time score increased, in the second test, was also significantly greater for the experimental group ($p < .001$).

($p. < .05$). The experimental group was also poorer on all the other problems, but the difference approached significance only in the case of anagrams ($p < .10$). No significant difference between the groups was obtained on tests given outside the cubicle.

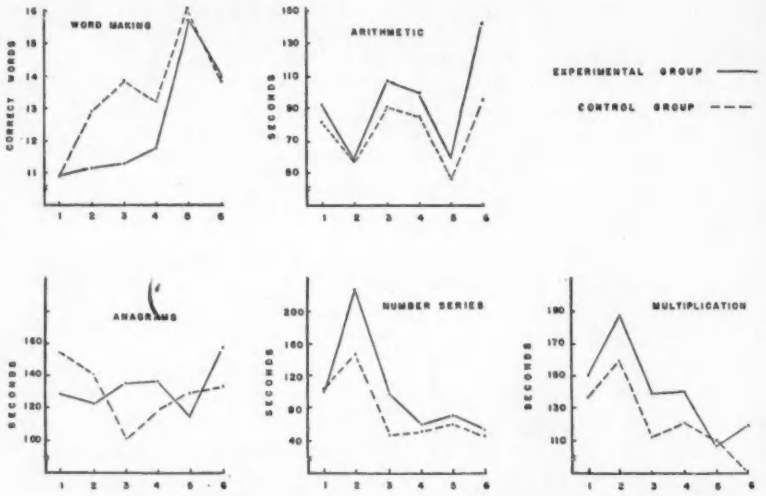


FIGURE 1. Scores for the experimental and control subjects on the five types of problem in cubicle battery A. Numbers along the abscissae indicate the test period. During test periods 2–4 the experimental subjects were in isolation.

When we compare the two groups on error scores, and on the number of requests for repetition of the problem, we find that during the isolation period the experimental group was again inferior to the control group. This is true for all tests and test periods during isolation, though the differences were significant only in the following cases: on the anagrams test the difference between the error scores of the two groups at the 24-hour period is significant at the 5 per cent level of confidence; on the word-making test the experimental subjects made significantly more errors at the first test period in the cubicle ($p < .05$) and at the last ($p < .02$). It is worth noting that they also made fewer words altogether (correct plus incorrect words) than did the control subjects (significantly fewer at the second test period in the cubicle, $p < .01$). On this same test, requests for repetition were significantly different at the first test period during isolation ($p < .05$).

There was virtually no difference in performance between the two groups on the digit span and analogies tests (cubicle battery, Part B). In associative learning the experimental subject did somewhat more poorly than the controls, but the difference did not approach significance ($p > .10$).

We must now consider the effects of propaganda. The *t*-test was used to find out whether the subjects had been affected by propaganda, and whether one group was more affected than the other. The results are summarized in Table III.

Both groups showed a significant change in attitude after listening to the propaganda, but the change was much greater for the experimental

TABLE III

MEAN TOTALS ON 5 ATTITUDE SCALES FOR 24 EXPERIMENTAL AND 35 CONTROL *Ss*, SHOWING INCREASED BELIEF IN PSYCHICAL PHENOMENA AFTER EXPOSURE TO POSITIVE PROPAGANDA*

| Scale | Possible score | Experimental | | Control | | $p$ |
|---|---|---|---|---|---|---|
| | | Pre | Post | Pre | Post | |
| Attitude | 168 | 68.7 | 106.5 | 73.4 | 91.5 | .02 |
| Interest | 45 | 23.5 | 31.0 | 26.9 | 29.6 | .01 |
| Importance | 20 | 9.0 | 13.0 | 9.4 | 11.3 | .01 |

*Scores showing increased interest and estimate of importance are also shown. $p$-values are based on a comparison of the mean difference scores (between the pre- and post-propaganda tests) for the two groups.

subjects than it was for the controls. On the total questionnaire (all scales) the difference between the groups was significant ($p < .02$). If we consider the various sections, we find that the differences are most marked on the sections dealing with ghosts, poltergeists, and clairvoyance, where the levels of confidence are beyond the 3, 5, and 7 per cent respectively. In the sections dealing with telepathy and psychical research, the differences, while in the same direction, do not approach significance. Table III also shows the results from two other subsidiary scales which measured the degree of interest which the subjects felt about the various topics, their previous familiarity with the topic, and the degree of importance which they felt that each topic had. Scores for the entire questionnaire are considered. It can be seen that although both groups are more interested in the subject after they have listened to the propaganda than they were before, and feel that it is of greater importance, the experimental subjects are affected to a greater degree than are the controls.

*During Isolation*

Of the 29 experimental subjects, nearly all reported inability to concentrate on any topic for long, lack of clarity in their thinking, or difficulty in organizing their thoughts. Most had originally planned to review their studies, solve problems, or "think about things" generally. But, after a few hours in isolation, they found that such efforts tended to be abortive. They reported that the disorganization in their thinking became more pronounced as the experimental period advanced, and described their thinking in the later stages with such words as sterile, garbled, disjointed, confused, ineffectual, shallow. They frequently experienced "blank periods" during which, though fully awake, they did not seem to be thinking of anything, and attributed this to the fact that they had "run out of topics" and "couldn't think of anything to think about." This, they felt, was one reason why the experimental situation became unpleasantly boring. The content of their thinking also changed as time went on: they reported that at first they did make some attempt to review studies, and to solve personal problems, but later resorted to reminiscence and idle day dreaming, making no attempt to control the direction of their thoughts.

There seemed to be some evidence of impaired judgment and loss of sense of proportion while the subjects were in isolation. Thus, one left the experimental situation highly pleased with his solution to a musical problem only to find that the phrase he had invented was, note for note, one that he already had in his manuscript. Another, towards the end of his stay, adopted the irrational procedure of "lying back and letting the problem sink in" in the hope that the answer would come to him "intuitively." Subjects frequently developed a childish sense of humour, and had exaggerated emotional reactions, becoming excessively irritated by small things and sometimes very annoyed with the experimenters. A few reported that they brooded about things unduly, dwelling on imaginary injustices. Some committed small deceptions (readily admitted later), and others became obsessive about trying to stay in the cubicle. Later, the subjects were frequently surprised by the way they felt during the isolation period, and said they had "magnified things," or had been "irritated out of all proportion."

*After Isolation*

The subjects' condition did not return to normal immediately after they emerged from isolation. That their behaviour was still impaired is evident from their test performance during the first 2½ hours after isolation (already described); it is confirmed by observation and their

own reports. They appeared dazed and somewhat unresponsive. Sometimes, during the tests, their attention would wander off the problem in hand, and they would sit staring into space. They commonly reported feeling remote and confused, and that they had difficulty in concentrating. Thus, one subject complained that he "had to relearn concentration"; some said they had to make a deliberate effort to concentrate on, and grasp, anything to which their attention was directed. Comments such as "I felt 'wooden' during testing," "impulses had difficulty getting to the brain," "my brain was fuzzy and couldn't grasp—it wouldn't take things in," "the patterns [of the Kohs' blocks] didn't sink into my mind—I had to *get* my mind to work," were common.

We gave no formal tests later than 2½ hours after emergence, other than final tests of the post-cubicle battery on the third and fourth days. Some subjects, however, indicated that they noticed some effects for two days after the experimental session ended, and reported a generalized lack of interest and loss of motivation to study. In addition, some reported behaviour of an "absent-minded" type. Two had difficulty driving their cars, one double-exposed his camera, another took a shower while partially clothed, and another forgot to bring his books to college. Several of our early subjects reported that they tended to plunge across the street without noticing or looking for traffic, and we had subsequently to warn subjects to be careful about this. In addition, almost all subjects reported an increased irritability during this period which made it difficult for them to get on with their associates.

The cubicle experience may have made the subjects unusually analytical of their own behaviour, and suggestion probably played a role, but it was clear that they regarded the incidents reported as unusual; moreover, most subjects reported the same symptoms, and were able to say when they disappeared.

## Discussion

The results, then, indicate that perceptual isolation produces a decline in intellectual ability. However, several puzzling results are evident. The difference between the groups did not increase as time went on, so that it seems as if there was no progressive deterioration. There are a number of possibilities why this result was obtained. One is that the cognitive abilities measured by the tests were affected early during isolation, but that there is no marked decline in performance between 12 and 48 hours of isolation, when the last test of the cubicle battery was given.

Had we given another set of tests at 72 hours, say, the difference between the groups might have been greater than it was at 48 hours.

Another possible reason is that each test session of the cubicle battery took up to two hours, and it became apparent that the control subjects regarded the tests as an unpleasant chore. Their motivation began to flag after about the second session. It is therefore possible that the control subjects' performance may also have been declining, so that the difference between the groups remained approximately constant.

When we consider the section of the experiment dealing with propaganda, we find that only 4 of the 35 control subjects wanted to hear the records more than once, in spite of the fact that they were paid by the hour to listen. On the other hand, 16 of the 24 experimental subjects asked for repetitions. This might lead the reader to believe that the results obtained were due solely to the fact that the experimental subjects spent a longer time listening to the propaganda than did the controls. Even if this were so, the results could still be attributed to isolation since one of the effects of isolating the individual is that he will listen to material which he would normally avoid, and listen to it often. Presumably, he becomes so bored that any form of stimulation is better than nothing at all. It happens, however, that there does not seem to be any relationship between the number of times that the subject listened to the records and the degree to which his attitudes were changed. No significant correlation can be detected.

To sum up, the results indicate that the experimental conditions produced some deterioration in performance on some tests of cognitive ability while the subjects were in isolation and after they had emerged. Subjects reported that their minds wandered, and that they were no longer able to find anything to think about. It seems too, that they became abnormally preoccupied with whatever patterned stimulation they did receive. These results may be attributed to some general disorganization of brain function which is also involved in the hallucinatory activity, disturbances of visual function and abnormal EEGs which occur under conditions of perceptual isolation (4).

## Summary

Twenty-nine male subjects were placed in isolation for as long as they would stay (usually 3 to 4 days). Two batteries of tests were given to them, before, during, and after isolation. In addition, they were subjected to propaganda during the isolation period. Twenty-seven control subjects, who were not isolated, were given the same tests and propaganda material.

The results indicate that the experimental subjects performed worse than the controls both during and after the isolation period on some tests, and that they were more susceptible to propaganda, though both groups showed a significant change in attitude.

## REFERENCES

1. Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, 8, 70–76.
2. Hebb, D. O. Man's frontal lobes. *Arch. Neurol. Psychiat.*, Chicago, 1945, 54, 10–24.
3. Heron, W. The pathology of boredom. *Sci. Amer.*, 1957, 196 (1), 52–56.
4. Heron, W., Doane, B. K., & Scott, T. H. Visual disturbances after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, 10, 13–18.

# CHANGES IN PERCEPTUAL FUNCTION AFTER ISOLATION[1]

B. K. DOANE,[2] WINSTON MAHATOO,[3] W. HERON, AND T. H. SCOTT[4]

*McGill University*

IN AN EARLIER preliminary report (4) we desribed some of the gross effects of perceptual isolation on visual perception, based on the subjective reports of three sophisticated subjects. In the present paper we provide data from a systematically administered battery of objective tests of visual perception given to thirteen subjects after four days of isolation, and data from a series of tests of somaesthetic perception and spatial orientation. In addition, we include results from a group of ambulatory subjects who underwent visual isolation, but were not restricted in any other way.

## METHOD

The method was that already described (7), with some slight changes in detail: the cubicle in which S was placed was now painted white, he had EEG leads attached throughout the experiment, and the isolation period of four days was agreed upon in advance. We also made certain physiological tests, and two tests of somaesthetic function, which briefly interrupted the isolation (visual perception, however, was occluded throughout). With the time necessary for eating and going to the toilet, S was out of the cubicle for about 4 hrs. out of 24.

The experimental group was made up of 17 male college students. Of these, 13 went through the cubicle procedure and 4 were ambulatory. Two cubicle Ss wore opaque masks, to test the effect of darkness on their hallucinatory activity, until an hour before the end of the experimental period, when they were fitted with the translucent mask worn by the others (this mask admitted light, but did not permit visual perception). The ambulatory Ss wore the translucent mask but not the gloves and cuffs which prevented tactual perception (7), and were not confined to the cubicle. They were worked with in pairs, and allowed to talk with each other, to listen to the radio, to go for walks and, in general, to engage in as much normal activity as was possible in the circumstances. These Ss were given the tests of visual perception only.

Tests of somaesthetic function and spatial orientation were given to 8 Ss, but because of a change of procedure, data are available for the 2-point limen on 5 Ss only.

Twenty normal control Ss were used. All were given the spatial and somaesthetic tests, and 13 were given the visual tests. The tests were given at the same time

[1]D.R.B. Project no. D 77–94–85–01.

[2]Now at the Faculty of Medicine, Dalhousie University.

[3]Now at the University of Montreal.

[4]Now at the University of Auckland, New Zealand.

intervals as for the experimental subjects, S coming into the laboratory at the agreed hour and being, of course, under no restriction at other times.

The tests were first given before S entered the experimental condition. Tests involving somaesthesis and orientation were repeated after 48 and 72 hrs. of isolation only (that is, not after the isolation period was over). Immediately upon coming out of restriction all Ss were asked to examine their visual surroundings and report on their appearance, their reports being recorded. The quantitative visual tests were then administered for the second time.

## Effects on Visual Perception

The tests of perception had to be completed in a short time since the major effects seem to wear off in an hour or two, and this limited our choice of tests to those that could be given quickly. Even so, after allowing about ten minutes for the qualitative description, the test battery took over an hour to administer, and it is likely that some of the later tests showed smaller differences than they would have if they had been given early in the series. The tests are listed below, together with the results, in the same order as given to the subject. Table I shows those in which the mean differences reached or approached significance.

1. No effect on *critical flicker frequency* was found.

2. There was an increase of *figural after-effect*, with the inspection and test figures of Köhler and Wallach's Figure 36 (6, p. 292) and the method of measurement described by them (p. 299). The measurement is of the degree of displacement of two test figures.

3. *Size constancy* was decreased. With each of 6 different-sized discs, presented at 3 ft., S made a comparison with a graduated series of 17 discs presented at 12 ft. He was instructed to "pick the far disc that looks the same as the near one," and his score was the difference in diameter between the test disc and the comparison disc that he chose as equal to it. The experimental Ss chose significantly larger comparison discs, showing a decrease in constancy; the control Ss showed no change (the slight difference in their mean scores was in the opposite direction).

4. *Visual acuity* was probably improved by the isolation procedure. A white card was presented at 10 ft., bearing a horizontal row of 14 black vertical lines spaced 1 in. apart. Each line was 1/64 in. wide and 3 in. in vertical extent. There was a gap in each line, decreasing from 3/32 in. at the left to 1/64 at the right. S had to say where the gap was in each line (bottom, middle or top). Three such cards were presented, making a total of 42 trials. Normal control Ss showed no change from first to second test; the experimental Ss improved from a mean score of 35.4 correct identifications, before isolation, to a mean of 36.4 afterwards. Though the result does not reach the usual level of statistical significance ($p < .10$), some weight may possibly be given to the result in view of the significant change of the 2-point limen (see below).

5. In the *phi-phenomenon* no changes were observed in the timing of the stimuli which gave rise to apparent movement.

6. No differences of *brightness contrast* effects were obtained when the method of Thurstone (9, p. 53) was used.

7. The *autokinetic effect* was increased. The subject was dark-adapted for 3 min.,

then seated 10 ft. from a point source of light in a dark room. Two minutes after onset of movement the illumination of the surroundings of the point source was gradually increased up to the point at which movement was abolished. Two such determinations were made. No difference was observed in time of onset of the movement, but movement for the experimental Ss persisted with a significantly higher level of surrounding brightness.

8. *Colour adaptation* increased. S looked with one eye through a 2-in. plain polaroid filter mounted in the front end of a box. Inside the box, 6 in. from the front, was a 2-in. yellow polaroid filter surrounded by a 4-in. white circular field. After fixating a point in the middle of the yellow disc for 90 sec., S closed his eyes for 10 sec. and then again looked at the fixation point while the yellow disc was rotated to reduce the saturation. Adaptation was measured by the number of degrees of rotation necessary to make the yellow disc match the surrounding white area.

9. *Shape constancy* may have decreased ($p < .10$). A white equilateral triangle with 3-in. sides was mounted on a grey card attached to supports so that it could be tilted at various angles from the vertical. The card was tilted at 5 different angles, and for each S was asked to select from a series of 15 triangles of various altitudes the one which seemed to match best the shape of the tilted triangle.

10. A test of *brightness constancy* showed no apparent effect of isolation.

11. *Necker cube* reversals showed no change of frequency.

12. *Movement after-images* were increased in duration. A 12-in. spiral, 12 ft. from S, was rotated at 2 r.p.s. for 30 sec. Ss report of after-movement was timed on 3 successive trials.

13. No effects on *tachistoscopic perception* were obtained. A series of black nonsense forms (outlines) were presented one at a time on a white screen for approximately 50 msec. A recognition method of testing was used.

In all these results, data for the ambulatory Ss are included with those for the more completely isolated Ss since they showed the same picture and, by increasing the *N*, improved the level of confidence for three of the above results (autokinetic movement, colour adaptation, and the after-image of movement.)

TABLE I

LEVELS OF CONFIDENCE FOR QUANTITATIVE TESTS SHOWING CHANGES OF VISUAL FUNCTION AFTER ISOLATION

| Test name | FAE* | Size constancy | Acuity | Auto-kinetic | Colour adaptation | Shape constancy | After-movement |
|---|---|---|---|---|---|---|---|
| Test number | 2 | 3 | 4 | 7 | 8 | 9 | 12 |
| Value of $p$ | .02 | .02 | .10 | .001 | .01 | .10 | .05 |

*Figural After-Effect

### *Qualitative Reports*

When the mask was first removed, all subjects in the cubicle group and three of the four ambulatory subjects reported gross visual disturbances, which usually disappeared in about half an hour. The effects described were similar to those reported by three subjects earlier (4); we are now able to give frequencies of occurrence for the various distortions in a larger group.

TABLE II

INCIDENCE OF VARIOUS DISTURBANCES OF VISUAL PERCEPTION IMMEDIATELY ON COMING OUT OF RESTRICTION: QUALITATIVE OBSERVATIONS BY 20 SUBJECTS

| Spontaneous movements* | Induced movements | Surface distortions | Linear distortions |
|---|---|---|---|
| 18 | 12 | 16 | 18 |

*For categories, see text.

Table II includes the results obtained earlier, and thus is based on an *N* of 20. It classifies the distortions under headings which may need some explanation. *Spontaneous movement* includes all apparent activity of the visual field when the observer (including his eyes) was still: shimmering or undulation of surfaces, and drifting, contraction, or expansion of objects. *Induced movement* refers to changes in the position of objects produced by head and eye movements. *Surface distortions* refers to cases in which plane surfaces were described as warped, concave, or convex, or as though folded along an axis in the medial optical plane, and those in which a convex swelling of the central part of the visual field was reported. *Linear distortions*: (*a*) with fixation on a point between two parallel vertical black lines on a grey background, about two-thirds of the subjects reported that the lines seemed to swell outward near the fixation point, and the remainder that the lines curved inward (toward the fixation point); (*b*) with fixation above and below a horizontal line, six of the twenty subjects reported that the ends of the line curved downward when fixation was below the line, upward when fixation was above. It should be emphasized that all these linear effects were pronounced: for example, the centres of the parallel lines (¼ in. wide, 3 in. apart, and 3 ft. from the observer) might according to the reports be displaced by one inch or more.

In addition to the distortions summarized in Table II, there were reports of exaggerated contrast, hypersaturation and luminosity of colours, pronounced positive and negative after-images, accentuated or diminished depth of perspective, and distortions of human faces. All of the effects described were obtained with both monocular and binocular vision, but were more marked binocularly.

*Factors Affecting Hallucination*

As an incidental aspect of the visual study, we were also able to obtain some further information on the factors determining the occurrence of the hallucinations which are described elsewhere (5). Of the eleven subjects who wore the translucent mask, and who were thus exposed continuously

to diffuse visual stimulation, eight developed hallucinations. Of the two subjects who wore the opaque masks, one developed hallucinations; but when they were given the translucent mask at the end of the isolation period, both had vivid hallucinations, those of the subject who had already had them becoming stronger.

This result suggested that exposure to diffuse light was a factor in the phenomenon. Accordingly, five subjects who were among the most persistent hallucinators when wearing the translucent mask were put in complete darkness. All reported an immediate increase in vividness of hallucinations; but within two hours there was a decrease of such experiences, three having no more hallucinations and two having them greatly decreased. Exposed again to the diffuse light, all five reported that the hallucinations returned to their original level of intensity.

A further fact of interest is that two of the four ambulatory subjects (with translucent mask) developed hallucinations; and further, that these began for one subject while he was being taken for a walk. These results indicate, first, that unpatterned sensory stimulation increases the probability of hallucinatory activity, as Vernon *et al.* have suggested (**10**), but is not necessary for its occurrence. Secondly, the hallucinations may be specific to restrictions in the particular sensory mode in which they occur —that is, the probability of visual hallucinations may not be affected by restriction, or lack of restriction, in other senses. It may be noted, however, that visual perception seemed somewhat more grossly disturbed in the cubicle subjects than in the ambulatory ones, which suggests that though the disturbance of function is greatest in the area in which restriction has occurred, there is also some spread of effect to other senses.

### Somaesthesis and Spatial Orientation

Two tests of somaesthetic function were made, a tactual form-discrimination test and measurement of the two-point limen. The two tests of orientation, one with paper and pencil and one in which the subject attempted to follow directions in moving about a bare room, might also be considered to be tests mostly of somaesthetic function, since they were done while still wearing the masks which prevented vision. The tests were given before entering the cubicle, after 48 hours of isolation, and after 72 hours. Eight subjects were tested, but data for the two-point limen are available for five subjects only because the procedure used at first was unsatisfactory. Also, one subject felt obliged to leave the experiment before the third test was due, so that the results for the third period are based on seven subjects only for form discrimination and orientation tests (Tables III and V), four subjects for the two-point limen (Table IV). In spite of the small *N*, however, significant results were obtained.

*Tactual form discrimination* was tested with ten figures made by fastening wire to the surface of square cards 6 × 6 in. Five were familiar geometric forms (circle, triangle, etc.), five were slight variations of such forms (ellipse, rectangle with one curved side). On the first (pre-isolation) test the subject was shown the forms. He then put on the mask and was required to identify the forms he had been shown, by tracing the outline of each with his finger for 10 sec. The forms were presented in variable order in the different tests. The second and third tests, after 48 hours and 72 hours of isolation, were made without showing him the forms first. Scores were based on errors (failures of identification).

TABLE III

MEAN ERROR SCORES ON FORM DISCRIMINATION, AT THREE TEST PERIODS, FOR 8 EXPERIMENTAL SUBJECTS (7 IN THE 72-HOUR TEST) AND 20 NORMAL CONTROL SUBJECTS

| Group | Test periods | | | $p$-values* ($U$ test) | |
|---|---|---|---|---|---|
| | Pre-isolation | 48 hrs. | 72 hrs. | 1–2 | 1–3 |
| Experimental | 2.06 | 3.30 | 2.70 | .001 | .02 |
| Control | 2.63 | 2.08 | 2.03 | | |

*$p$-values are based on each subject's change in score as between his first and second, and first and third test.

The results are summarized in Table III: control subjects showed some slight practice effect in the second and third test periods, experimental subjects deteriorated significantly. The statistical analysis was made in terms of difference scores, obtained by subtracting the first test score from the later score, for each subject.

*Two-point limens* were measured with a standard aesthesiometer on four loci: (1) tip of left index finger, (2) volar surface of left forearm, (3) inner surface of the right upper arm, three inches above the joint, and (4) the forehead one inch above the nasion. To ensure testing the same point on all trials, a line was drawn on the skin with indelible ink. One tip of the aesthesiometer was applied at the end of the line, the other at various points along its extent. The method of limits was used, with modifications, to limit the expenditure of time. At each test period, only one ascending and one descending series of determinations was made until the subject responded correctly three times. Single points were presented at random.

The results (Table IV) showed a decrease of two-point limen for the experimental group in three of the four loci, two of these at conventional levels of significance. The analysis again was based on difference scores, comparing the amount of *change* shown by experimental and by control

TABLE IV

MEAN VALUES FOR TWO-POINT LIMEN FOR 5 EXPERIMENTAL SUBJECTS (4 IN THE 72-HOUR TEST) AND 20 NORMAL CONTROL SUBJECTS

| Locus | Group | Test periods | | | $p$-values ($U$ test)* | |
|---|---|---|---|---|---|---|
| | | Pre-isolation | 48 hrs. | 72 hrs. | 1–2 | 1–3 |
| Finger | Experimental | 1.70 | 1.70 | 1.75 | NS | |
| | Control | 1.75 | 1.50 | 1.60 | | |
| Forearm | Experimental | 29.6 | 26.7 | 24.0 | .15 | .15 |
| | Control | 23.4 | 23.3 | 23.7 | | |
| Upper arm | Experimental | 29.1 | 21.9 | 23.8 | .002 | .05 |
| | Control | 32.8 | 32.8 | 32.4 | | |
| Forehead | Experimental | 19.8 | 16.9 | 19.2 | .02 | .02 |
| | Control | 9.2 | 9.2 | 9.2 | | |

*$p$-values as in Table III.

subjects: this is an important point in understanding the data for the forehead as shown in Table IV, since the limen of one of the experimental subjects was well outside the range for the others, and this gives the experimental group much higher means than the control group.

One peculiar phenomenon was noticed while the experimental subjects were being tested: they were sometimes uncertain whether they were being touched or not, and would frequently respond when no stimulus was being applied. All of them behaved in this way though not on the pre-isolation tests; only one control subject did so, on one occasion. It seems, then, that they were experiencing some form of "hallucination"; they felt that they were being touched by the aesthesiometer when this was actually not so.

*Spatial orientation* was tested in two ways, both requiring that the subject follow directions in making a series of movements without vision. The first was a paper-and-pencil test. The pencil held by the subject was placed at the starting point on a sheet of paper, and he was instructed to move left three inches, make a right turn and go two inches, another right turn and two inches, a left turn and three inches, and so on, with a total of five turns. He was instructed that each turn was to be a right angle, and was warned that at the end he would be required to draw a straight line back to his starting point. The scoring system was based on deviations of distances and angles from the correct ones. Two such tests were given.

The second way in which orientation was tested is the same in prin-

ciple, but here the subject was placed in an empty room and traced the patterns by walking, and his position and orientation were corrected after each individual response.

TABLE V

MEAN TOTAL ANGULAR DEVIATIONS IN SPATIAL ORIENTATION TESTS FOR 8 EXPERIMENTAL SUBJECTS (7 AT THE 72-HR. TEST) AND 20 NORMAL CONTROL

| Test | Group | Test periods | | | $p$-values ($U$ test)* | |
|---|---|---|---|---|---|---|
| | | Pre-isolation | 48 hrs. | 72 hrs. | 1–2 | 1–3 |
| Paper and pencil | Experimental | 40.6 | 140.0 | 56.4 | .05 | .05 |
| | Control | 58.8 | 57.0 | 53.3 | | |
| Walking | Experimental | 112.5 | 121.3 | 135.0 | .10 | .001 |
| | Control | 146.5 | 115.8 | 102.8 | | |

*$p$-values as in Table III

In the estimation of distance the experimental subjects did not differ from the control, but they were inferior in the judgment of angles and directions. Table V presents the test results. To these some qualitative observations may be added. In the walking test, when the subject was told to return to his starting point, he sometimes became quite disoriented, without even an approximate idea of the direction in which to move. This was not observed in the control group. Disorientation was also observed at other times. Before the subject entered the cubicle he was shown the layout of the washroom, and during the experiment the procedure for "toileting" was to escort him to the door and then let him find his way to the toilet and back to the door by himself. In the early stages of isolation the subject had no difficulty in doing so, but in later stages he fairly often became lost and would have to call the experimenter to help him find his way out.

## SUMMARY AND COMMENT

Quantitative data are provided concerning the changes in sensory and perceptual processes, both visual and somaesthetic, that result from the isolation procedure. Some further information is also provided about the factors affecting the occurrence of hallucinations.

There are certain general areas of perception which seem to be affected by the isolation procedure. Results from the visual tests indicate that the most prominent effects are a decrease in the constancies and an increase in the after-effects of stimulation (figural after-effect, colour

adaptation, and the after-image of movement). Size constancy is markedly reduced, and shape constancy probably reduced; to these may be added the repeated observation of subjects on first emerging from isolation that the position of objects in the visual field was unstable, moving when the subject moved. As Gibson (2) has pointed out, the stability of objects in the visual world must be regarded as a constancy phenomenon.

It seems possible, further, that visual acuity was improved by the isolation procedure. The 10 per cent level of confidence, statistically, can of course justify no final conclusion, and the apparent support from the statistically significant improvement in the two-point limen must also be regarded cautiously since it is based on data from five subjects only (though in comparison with twenty control subjects). However, it is worth noting the similarity of this result to the reports of Haber (3) and Teuber, Krieger, and Bender (8) of decrease in the two-point limen in the limb stumps of amputees, results which they interpret in terms of a central reorganization due to reduced sensory input.

Finally, we may note the possibility that the experimental subjects' poorer performance in the tactual form-discrimination and spatial orientation tests was in fact caused by visual dysfunction. Some subjects remarked that they had lost their ability to visualize the external world. This made it impossible for them to form a mental picture of the route that had been followed in the orientation test, and thus to locate the starting point. One subject who got lost in the confines of the washroom reported that "One reason why I can't find my way is that when I try to visualize where I am everything seems to be expanding and contracting and waving about." Similar remarks were made by others, and it is relevant to recall that such remarks were typically made by subjects when they were first describing their hallucinatory activity, or the first appearance of the external world when they first emerged from isolation.

These results emphasize again the profound degree of disturbance that is produced by the isolation procedure as observed in this laboratory and elsewhere. Hallucinations of extreme vividness, impairment of thought processes, sensory and perceptual changes, together with significant changes in the EEG, all testify to the widespread effect on central neural function that is induced simply by limiting the normal variation of sensory stimulation.

There is as yet no sign of a satisfactory general explanation of these phenomena, and it is clearly impossible to give one which is plausible and detailed. But one might guess that the functional de-afferentation of the isolation conditions may cause parts of the central nervous system to become hyperexcitable. Evarts (1) has recently shown that such phenomena may take place in the visual system of the cat. Using

chronically implanted electrodes, he has demonstrated that the cortical response to electrical stimulation of the optic pathways has a shorter recovery cycle after the animal has worn opaque contact lenses for a few weeks.

In our subjects there was no diminution of total sensory stimulation but, since sensory systems respond most actively to change of stimulation, it is possible that the lack of a varied input results in an inactivity of pathways at some higher levels of the central nervous system. If these pathways consequently became sensitized, it might account for increased figural after-effect, autokinetic movement, colour adaptation, and so on. It would also fit in with the least expected feature of the present experiment, the apparent increase in sensory acuity, visual and somaesthetic (the only example that we have found of an improvement of function after isolation). On the other hand, it would interfere with the integration of the complex patterns of neural activity which must be involved in more complicated tasks such as problem solving and spatial orientation, the constancies being also affected. Finally, it is easy to see how hyperexcitability of parts of the visual system might be related to the occurrence of hallucinations. Until we have more detailed information, however, such ideas must remain speculative.

## REFERENCES


1. Evarts, E. The physiological basis of hallucinations. In Jolyon West (ed.), *Hallucinations: A symposium* (in press).
2. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.*, 1954, **61**, 304–314.
3. Haber, William B. Effects of loss of limb on sensory function. *J. Psychol.*, 1955, **40**, 115–123.
4. Heron, W., Doane, B. K., & Scott, T. H. Visual disturbances after prolonged perceptual isolation. *Canad. J. Psychol.*, 1954, **8**, 70–76.
5. Heron, W., Doane, B. K., & Scott, T. H. Characteristics of hallucinations during perceptual isolation. In Jolyon West (ed.), *Hallucinations: A symposium.* (In press)
6. Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
7. Scott, T. H., Bexton, W. H., Heron, W. & Doane, B. K. Cognitive effects of perceptual isolation. *Canad. J. Psychol.*, 1959, **13**, 200–209.
8. Teuber, H. L., Krieger, H. P. & Bender, M. B. Reorganization of sensory function in amputation stumps: Two-point discrimination. *Fed. Proc.*, 1949, **8**, 156.
9. Thurstone, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.
10. Vernon, J., McGill, T. E., & Schiffman, H. Visual hallucinations during perceptual isolation. *Canad. J. Psychol.*, 1958, **12**, 31–34.

## BOOK REVIEWS

*Contributions to Modern Psychology: Selected Readings in General Psychology.* By DON E. DULANY, JR., RUSSELL L. DEVALOIS, DAVID C. BEARDSLEE, and MARIAN R. WINTERBOTTOM. Toronto: Oxford University Press, 1958. Pp. x, 398. $3.25.

I CONFESS to an old-fashioned prejudice against books of selections, yet in these days of large classes and overcrowded libraries I am not unresponsive to their practical advantages. Of the several similar productions which have recently appeared, this seems to me one of the very best. The format and typography are worthy of the distinguished press whose imprint it bears. The contents have been so discreetly and unobtrusively edited (not virtually rewritten in simplified form as is the case with at least one of the others) that the freshness and individuality of the work of each contributor are preserved. Even the lists of references at the ends of the articles have been retained. Yet the beginner will hardly find anything more difficult of comprehension than the matter he encounters in his introductory textbook, unless perhaps it be E. G. Boring's high-spirited but at times baffling "When is Human Behavior Predetermined?" The student of today should be grateful that the path is made smoother for him by so engaging and informative a companion to his first psychological studies. The papers are extremely readable and in general they fulfil adequately the expressed intention of the editors to tell the student more about "how the psychologist as investigator works" than is attempted in the ordinary textbooks. If there is to be criticism, it must be that while many of the papers deserve the editors' adjective "classical" and all are interesting, not all are on quite the same level of scientific importance or methodological significance. The question might also be asked whether in the long run the student may not gain more from the closer study of a smaller number of topics, but this is a question which might equally be raised of the majority of those widely used textbooks whose study these readings are designed to accompany and complement. Clearly it is impossible to produce a book of this kind which will give equal pleasure to everyone, but I believe that this selection successfully represents the kind of things that are most likely to be acceptable to the majority of instructors of introductory classes. Of course, it is assumed that the reader will have the integrating guidance of an instructor and a textbook, otherwise his impression of the book, and of psychology, may well be that of the old lady who, given a dictionary as a birthday present, remarked after going conscientiously through it that it was "quite fascinating to read but dreadfully disconnected."

The book contains forty-two selections varying in length from two to

twenty-eight pages, and averaging about nine and a half. Three were published as recently as 1957, and altogether twenty-five are from the 1950's. Of the remainder, ten date from the 1940's, five from the 1930's, one (Freud) is from 1908, and one (from Frazer's *The Golden Bough*), though here dated 1949, originally appeared in 1890. Length and recency of publication, however, are not directly proportionate to theoretical importance. Five are drawn from books, though four of these are themselves collections of papers, three by various hands. Of the journals, the *Journal of Experimental Psychology* is represented five times, the *Journal of Abnormal and Social Psychology* four times, the *American Psychologist*, *Science* and the *Scientific American* each three times. Fourteen other journals are drawn from, all but four of them American. The *Canadian Journal of Psychology* is represented by Bexton, Heron, and Scott's paper on "Effects of Decreased Variation in the Sensory Environment", and the book opens with D. O. Hebb's "Heredity and Environment in Behavior" reprinted from the *British Journal of Animal Behaviour*. The names of many of the contributors are household words. The selections are arranged under eight headings (the number of papers in each is indicated in parentheses): Development (4), Intelligence and Individual Differences (4), Perception (7), Learning and Memory (6), Thinking and Imagination (4), Motivation and Emotion (8), Personality (5), Personality Disorders and Their Treatment (4). While there is no separate section for Social Behaviour, several of the papers could quite properly have been grouped under this heading had it been desired. Comparison with other books of roughly the same sort shows that with the forty-six selections in the second edition of Hartley and Hartley's *Outside Readings in Psychology* (1957), the seventy in Dennis' *Readings in General Psychology* (1949), and the ten in Halmos and Iliffe's *Readings in General Psychology* (1958), there is no duplication. Three of the papers, however, also appear (much abridged and rather heavily edited, it is true) among the thirty-three in McGuigan and Calvin's *Current Studies in Psychology* (1958).

F. Hilton Page

*University of King's College*

*Annual Review of Psychology, Vol. 10.* Palo Alto, Calif.: Annual Reviews, Inc., 1959. Pp. vi, 520. $7.50.

Readers of the tenth edition of the *Annual Review of Psychology* will be overwhelmed by the enormous literature which is here surveyed—close to 2,500 articles written by some 2,600 authors. These articles examine such a range of topics and postulate for their evaluation such a varied technical competence that the reader winds up with a deep sense of

humility, if not utter frustration. It is easy for him to grasp, by analogy, the concept of an expanding universe!

These surveys cover but a small fraction of the literature, and a rigorous selection had to be made. All readers must admire the erudition and tenacious industry of the reviewers; many will not feel happy about the choice of some articles and topics or the allocation of space to the various areas.

Changes in this tenth *Review* are the omission of chapters on Perception, Industrial and Engineering Psychology, and (owing to the illness of one reviewer) Comparative Psychology. Included is a much needed chapter on Problem Solving and Thinking, by Gagné, who in spite of some pedantic attempts at humour, evaluates his material in a challenging fashion. For the first time since Volume 3 there is a chapter on Motivation, by Cofer, who restricts himself to an exposition of what he calls his first cluster of topics in this area, mainly physiological and comparative. While he discusses some intriguing findings in these areas, it seems a sad revelation of the all too prevalent trend of many of our brightest minds in psychology, chiefly in the learning and motivational fields, to focus their energies too exclusively on behavioural studies of animals, while disregarding almost completely the all-important study of man in his social environment.

This reviewer was particularly impressed by the chapter written by White, with his smooth evaluation of Abnormalities of Behavior and his lucid analysis of cycles of discovery in research, by Loevinger's terse and acerb chapter on Assessment, by Gebhard's chapter on Vision, and by Rosvold's incisive survey of physiological studies.

On the debit side there is the chapter by Mintz who strives to wring significance out of many trivial Russian studies. Most of the 30 pages given to this article could profitably be allocated to other topics, especially when one notes some 12 textual pages allotted to Statistical Methods and 17 to Abnormalities of Behavior. Furthermore, while the survey of Russian literature is of value to American and Canadian psychologists who are too apt to have a blind spot for the literature of other nations, one deplores the fact that few or no articles in the German, French, or Italian literature appear in many reviews. A meritorious exception is the chapter by Gebhard who quotes some 50 foreign articles on vision. It is distressing to note that only a score or so of articles from the *Canadian Journal of Psychology* are quoted.

Most reviewers are optimistic about the future of research in Psychology but nearly all feel that fresh, bold approaches, with sharpened basic concepts and a tightened methodology, are sadly needed.

B. R. PHILIP

*Assumption University of Windsor*

*Psychopathology, a Source Book.* Edited by CHARLES F. REED, IRVING E. ALEXANDER, & SILVAN S. TOMKINS. Cambridge, Mass.: Harvard University Press (Toronto: S. J. Reginald Saunders & Company, Ltd.), 1958. Pp. xii, 804. $13.75.

THIS BOOK IS A COLLECTION of 46 separate papers that originally appeared in a variety of psychological and psychiatric journals, mostly between 1952 and 1957. The papers are categorized under five headings: psychopathology and early experience, psychosomatic disorders and neurosis, schizophrenic psychoses, somatic factors in psychopathology, and psychopathology and the social context. A little over half of the contributions report original data, about ten are reviews, and about the same number theoretical. The papers range from reports of naturalistic observations to those of strictly experimental investigations, and cover a wide variety of methods, including ratings, follow-up assessments, retrospective analyses, the Rorschach test, and muscle tension recording. Though the editors selected the papers primarily on the basis of their research significance rather than their representativeness, they seem to have ended up with a volume that gives a fair picture of psychopathological research today.

As might be expected in a collection of this type, some papers are better than others. With only a couple of exceptions, the theoretical papers seem to be loosely organized and present inadequately sharpened ideas; this is particularly true of the papers written within the psychoanalytic framework. In general, those empirical investigations and reviews that are written with minimum reference to complex (and often vague) theoretical frameworks stand out as the most interesting ones. There are several such stimulating papers, including one on the psychological factors in attacks of asthma, one on hypnotic deafness, one on conditioned autonomic responses and anxiety, two on the effects of psychotomimetic drugs, and a few on the influence of early experience. In short, there is little doubt that the editors of this book have succeeded in their stated aim of providing a number of interesting papers to supplement textbooks for courses in abnormal psychology and psychiatry. However, in this reviewer's opinion, the deletion of several of the papers would have improved the selection as a whole. These include a paper on the concept of ego disturbance and ego support, the discussion of Fairbairn's theory of schizoid reactions, and a good but irrelevant article on the University of California oath controversy. There is some chaff with the grain; *Psychopathology* will be best used when it is used selectively.

This last point leads me to a more general question. Is a book of readings the most effective and inexpensive (this book costs $13.75) way of making supplementary material available to the student? I do not pretend to know the answer to this question, but it is my impression that

every instructor, considering his adopted textbook, likes to select his own list of supplementary readings. Books of readings are useful because they substitute for journals that are not readily available to most students, but such books can hardly be expected to meet exactly the needs of different instructors, each teaching abnormal psychology in his own way. Therefore, books of collected papers are only half successful in their aim. Departmental reprint libraries may be one answer to the problem. Again, in this day of duplicating and off-set machines, perhaps a scheme can be devised that would enable each instructor to prepare his own supplementary reading material (with due regard to the copyrights of the various authors). However, until such a flexible scheme is developed and becomes widespread, books of readings will continue to be useful to students and instructors alike. I have no doubt that *Psychopathology*, which has been put together by three experienced teachers, will be an aid in fulfilling the needs of many instructors and stimulating students.

DALBIR BINDRA

*McGill University*

*The Psychology of Interpersonal Relations*. By FRITZ HEIDER. New York: John Wiley & Sons, Inc., 1958. Pp. ix, 322. $6.25.

TODAY, WHEN IT APPEARS that all psychology is learning and all learning is reinforcement, many sections of our traditional subject-matter are in danger of neglect. One of these is man's everyday social behaviour—how he perceives people and recognizes their aims, motives, and sentiments towards himself. Asch and others have done something to keep these topics alive, and now Heider has published a book on the subject, one that embodies almost a lifetime of study and reflection.

Psychology, says Heider, is unique among the sciences in that common sense—the "intuitive" knowledge that all normal people acquire—is more complete and penetrating here than anywhere else. Deprived of what physical science has discovered, our society would relapse at once to a primitive level, whereas removal of all knowledge of scientific psychology would make practically no difference in our daily handling of human relations. This sobering fact reflects not only the backwardness of social psychology, but also the high development of that empirical knowledge which social life itself provides. "Common sense psychology," usually derided by introductory course instructors, is shown by Heider to be extensive, detailed, well organized, and surprisingly powerful. And why not? It results, after all, from the most prolonged and intensive learning that the human being undergoes, and it is also (as this book makes clear)

constructed and ordered by methods basically identical with those of science itself.

How do we make sense of the multifarious acts directed by others towards, against, or with regard to ourselves? Heider answers that "man grasps reality, and can predict and control it, by referring transient and variable behavior and events to relatively unchanging underlying conditions." This is sound Newtonian science. And, when the behaviour is human, we typically refer it to the motives, sentiments, beliefs, and personality traits of others. These are the constructs by which we integrate a bewildering mass of data in economical terms. When we see a new scratch on the car in the driveway we never merely register and respond to the visual stimuli as such. Instead, we immediately push deeper with such questions as: Who did it? Was it an accident? If not, was the doer trying to injure me, or merely feeling destructive? Until we have answers of some sort, and can see the occurrence as an instance of some empirical psychological generalization, we are not content. Thus, it is only in the laboratory that human beings behave like S-R machines; elsewhere they behave like theoretical scientists.

This universal search for causal connections is, of course, the prototype of "research." All that science does is to reduce the probability of error in our thinking, and to achieve greater generality by replacing our naïve constructs (people, intentions, feelings, etc.) with concepts such as drive, learning, and habit. These are defined at a high level of abstraction and are thus unavailable to the layman. For the concrete predictions on which he depends he must still rely on common sense; science may ultimately improve it, but can never replace it.

Heider has long viewed common sense psychology with the respect it deserves. He here attempts to lay bare its systematic character and some of its assumptions and structure by arranging samples of cognitive behaviour (inferences, etc.) in such a way as to expose the implicit rules operating, the "grammar" involved. He is modest about his achievement, regarding it only as a first, tentative effort to make the intuitive theories explicit. But he has done a remarkable job. The sheer number of situations dealt with is impressive, and they are analysed with great skill, not in terms of depth psychology, Freudian or Behaviorist, but at the phenomenal level where we ourselves operate.

The author has an immensely fertile mind, and there are points where his listing of all conceivable variants of a situation becomes tedious. But several of the chapters (notably that on sentiment) are astonishingly original and penetrating, and there is hardly a page without some revealing insight. One recalls his subtle treatments of the difference between personal and impersonal causation, of the feeling of "ought," of

revenge and retribution; and his brilliant demonstration of the quantitative "balance" that functions when sentiment and action conflict. Here the analysis goes beyond the phenomenal field to a novel kind of information theory.

As might be expected, Heider finds his observations foreshadowed and supported in philosophy and literature from Aristotle to Sartre. But the writers are more than outweighed by psychologists, not only Gestaltists, but Floyd Allport, J. J. Gibson, Skinner, Bruner, Dollard, Hebb, and dozens more. Only a rare psychologist could have written this book, one whose professional training has not erased his understanding of life and literature.

Heider's principles are so far of a rather low order of generality; whether they can be developed to full theoretical status without abandoning the phenomenal level may be doubted. But at least he has shown us how we manage the interpersonal encounters in which we are daily and hourly engaged, how we interpret them and evaluate them. And if this is not psychology, what is it?

J. D. Ketchum

*University of Toronto*

*Personality and Temperament.* By Solomon Diamond. New York: Harper & Brothers, 1957. Pp. vii, 463. $6.00.

This is not just another book on personality and temperament. The familiar facts and theories are all represented, but the author's thoroughgoing comparative-experimental approach makes this a welcome contribution to an area still dominated by disjointed presentations of irreconcilable viewpoints.

For the purpose of this study, temperament is defined "as including those aspects of individuality which depend on the ease of arousal of innate patterns of response." Its main dimensions are said to be fear, aggression, inhibitory control, and affiliative tendencies. For each of these, some animal-experimental evidence is cited from various levels of the phylogenetic scale between rodents and man. From this, the author derives the generalization that "the characteristics of the species establish a pattern which cannot be violated. The environment modifies, but it does not create, the pattern of individual life."

Refreshing though this viewpoint is in contemporary personality study, it does in this book lead to some dogmatic assertions with regard to such matters as critical periods in the temperamental development of children. Because effective cortical inhibitory processes are absent in

the first few months of life, the point is made that this normal instability of the infant could predispose to anxiety or generalized fearfulness if frequently precipitated in this period. A similarly unconvincing argument is advanced with respect to the acquisition of dependent-affiliative behaviour in early infancy.

Unfortunately, the otherwise excellent discussion of personality development at the human level is interrupted by two, primarily technique-oriented, chapters which would have been more appropriately placed elsewhere. The first deals with physiological theories of temperament and the other with factorial methods. In both the selection of representative material was evidently guided by the author's primary affinity for counselling rather than for clinical psychology. Thus, Sheldon and Cattell and not Kretschmer or Eysenck occupy the limelight in these chapters which offer a lucid introduction to their respective methodologies.

Personality, here defined as those aspects of individuality which arise from distinctively human capacities, is dealt with in the second half of the book. In the reviewer's opinion, the best feature of this section is the prominence given to the concept of "feeling habits" acquired by emotional learning. In this connection, the reminder that the important early training which influences later personality is not the learning of motor responses but the learning of feeling habits should appeal to the educationally oriented reader.

The most speculative aspects of the subject-matter, notably dynamic theory, are dealt with under the headings of maturity level, formation and defence of the self concept, interests, and, strangely enough, wisdom and laughter. There are also two chapters on cognitive individuality. These make the important point that general intelligence and special mental abilities are also dimensions of the personality sphere. Beyond that, however, these are the weakest chapters in the book constituting at times little more than a *pot-pourri* of divergent themes ranging from isomorphism and brain models to dream symbolism and projective techniques.

As the jacket cover asserts, the book is likely to stimulate independent thought about problems of human behaviour. It will do this most effectively at the undergraduate level where it will have few rivals with respect to the reconciliation it achieves between biological and social approaches to the study of personality.

ERNEST G. POSER

*McGill University*